Dynamic Multivariate Loss and Risk Assessment of Process Facilities

by

© Seyed Javad Hashemi

A thesis submitted to the

School of Graduate Studies

in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

Faculty of Engineering and Applied Science

Memorial University of Newfoundland

October 2016

St. John’s Newfoundland

Dedicated to
my wife and my parents
for their patience and faith

ABSTRACT

Dynamic risk assessments (DRA) are the next generation of risk estimation approaches that help to enable safer operations of complex process systems in changing environments. By incorporating new evidences from systems in the risk assessment process, the DRA techniques ensure estimation of current risk. This thesis investigates the existing knowledge and technological challenges associated with dynamic risk assessment and proposes new methods to improve effective implementation of DRA techniques.

Risk is defined as the combination of three attributes: what can go wrong, how bad could it be, and how often might it happen. This research evaluates the limitations of the methodologies that have been developed to answer the latter two questions. Loss functions are used in this work to estimate and model operational loss in process facilities. The application of loss functions provides the following advantages: (i) the stochastic nature of losses is taken into account; and (ii) the estimation of the operational loss in process facilities due to the deviation of key process characteristics (KPC) is conducted. Models to estimate reputational loss and significant elements of business interruption loss, which are usually ignored in the literature, are also provided. This research also presents a methodology to develop multivariate loss functions to measure the operational loss of multivariate process systems. For this purpose, copula functions are used to link the univariate loss functions and develop the multivariate loss functions. Copula functions are also used to address the existing challenge of loss aggregation for multiple-loss scenarios.

Regarding the dynamic estimation of the probability of abnormal events, the Bayesian Network (BN) has usually been used in the literature. However, integrated safety analysis

of hazardous process facilities calls for an understanding of both stochastic and topological dependencies, going beyond traditional BN analysis to study cause-effect relationships among major risk factors. This work presents a novel model based on the Copula Bayesian Network (CBN) for multivariate safety analysis of process systems, which addresses the main shortcomings of traditional BNs. The proposed CBN model offers great flexibility in probabilistic analysis of individual risk factors while considering their uncertainty and complex stochastic dependence.

The research outcomes provide advanced methods for critical operations, such as the offshore operations in harsh environments, to be used in continuous improvement of processes and real-time risk estimation. Application of the proposed dynamic risk assessment framework, along with a proper safety culture, enhances the day-to-day risk-informed decision making process by constantly monitoring, evaluating and improving the process safety performance.

ACKNOWLEDGMENTS

My deepest gratitude is to my supervisors, Dr. Faisal Khan and Dr. Salim Ahmed, for holding me to a high research standard and teaching me how to conduct succesful research. I have been fortunate to have Dr. Faisal Khan as an advisor who taught me how to question thoughts and express ideas. His patience, support, and prompt feedback helped me overcome many challenges and finish this dissertation. I am especially thankful to Dr. Faisal Khan for his friendship and for sharing his vast experience and knowledge over the past four years.

Dr. Salim Ahmed has been always there to listen, support and give advice. I am deeply grateful to him for the helpful discussions, insightful comments and constructive criticisms at different stages of my research, which were thought-provoking and helped me focus my ideas. I am indebted to him for his continuous encouragement and guidance.

I am also grateful to Dr. Yan Zhang, a member of my supervisory committee, for her encouragement and inspirational questions. I am indebted to the members of the Centre for Risk, Integrity and Safety Engineering (C-RISE) with whom I have interacted during the course of my PhD study.

I would like to acknowledge the financial support through the Ocean Industries Student Research Award (OISRA) provided by Research and Development Corporation (RDC) Newfoundland and Labrador as well as the financial support provided by Memorial University of Newfoundland, Natural Sciences and Engineering Research Council (NSERC) of Canada, the Atlantic Canada Opportunities Agency (ACOA), and the Vale Research Chair Grant.

Most importantly, none of this would have been possible without the love and patience of my family. My beloved wife and my immediate family have been a constant source of love, concern, support and strength all these years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS AND SYMBOLS

Acronyms

AIC	Akaike's information criterion
AL	Asset loss
BBL	Barrel
BIC	Bayesian information criterion
BN	Bayesian Network
CBN	Copula Bayesian Network
CCPS	Center for Chemical Process Safety
CDF	Cumulative distribution function
CML	Canonical maximum likelihood
COF	Consequences of failure
CPT	Conditional probability tables
CSTR	Continuous stirred tank reactor
DAG	Directed acyclic graph
DCBN	Dynamic Copula Bayesian Network
DF	Deviation frequency
DP	Differential pressure
DR	Demand rate
DRA	Dynamic risk assessment
DRAF	Demand rate adjustment factor
ECC	Environmental Cleanup Cost
EDF	Empirical distribution function
EE	End event
EML	Estimated maximum loss
ESD	Emergency shutdown device
FMEA	Failure modes and effects analysis

FTA	Fault tree analysis
HA	High alarm
HAZOP	Hazard and operability study
HEP	Human error probability
HHA	High-high alarm
HTE	High temperature event
IBLF	Inverted Beta loss function
IE	Initiating event
IFM	Inference functions for margins
INLF	Inverted normal loss function
IPLF	Inverted probability loss function
KPC	Key process characteristic
KL	Kullback-Leibler (KL) distance
LF	Loss functions
LIR	Loss inversion ratio
LLA	Low-low alarm
LOC	Loss of containment
LOPA	Layer of protection analysis
LSL	Lower specification limit
LTE	Low temperature event
MINLF	Modified inverted normal loss function
MC	Monte Carlo
ML	Maximum log-likelihood
MPD	Managed pressure drilling
MVD	Multivariate density
PCA	Principal component analysis
PDF	Probability distribution function
PFD	Probability of failure on demand
PL	Production loss

PERT	Program Evaluation Review Technique
PHA	Preliminary hazard analysis
POF	Probability of failure
ppg	Pounds per gallon
PRD	Pressure relief device
psi	Pounds per square inch
PSM	Process safety management
QLF	Quadratic loss function
SB	Safety barrier
SEM	Structural equation models
SIF	Safety instrumented function
SIS	Safety instrumented systems
StD	Standard deviation
UF	Uncertainty factor
USL	Upper specification limit
VSL	Value of statistical life

English letters

A	Root cause node
B	Constant in the QLF
b	Bootstrap independent data unit
C	Copula function
c	Copula density function
C_{fo}^{dt}	All fixed operating expenses which will NOT be incurred during process downtime in $/day
C_d	Expected costs to fix the damage and restore the business in $
C_L	Constant in the IBLF
D	Dataset
DF	Deviation frequency

DP	Differential pressure
DR	Demand rate
d	Dimension of a dataset
DP_T	Target differential pressure
E	Event node
ECD	Equivalent circulating density
EML	Estimated maximum loss
$E\{l(\cdot)\}$	Expected loss value
$E^*\{l(\cdot)\}$	Unit expected operational loss value
$f(x)$	Probability density function
$F(x)$	Cumulative distribution function
FR	Failure rate
G	Directed acyclic graph
H	Multivariate distribution
h	Drilling height
IC	Insurance coverage
J	Number of runs in Monte Carlo simulations
k_g	Shape parameter of Gamma distribution
$L(\cdot)$	Loss value
$L(x,T)$	Loss for deviations of x, given target T
ℓ	Likelihood function
L_0	Class 0 operational loss
L_I	Class I operational loss
L_{II}	Class II: business loss
L_{III}	Class III: event loss
L_ε	Residual losses/compensations

$L_{\text{Scenario } i}$	The cumulative loss of Scenario *i*
l_0^O	Allowable operational loss
l_I^O	Non-acceptable operational loss
l_{II}^{BI}	Business interruption loss
l_{II}^R , L_R	Reputational loss
l_{III}^A	Asset loss
l_{III}^{HH}	Human health loss
l_{III}^{EC}	Environmental cleanup loss
l_{profit}^{dt}	Profit loss due to lost production over process downtime in $/day
l_{profit}^{rp}	Profit loss over recovery period in $/day
pa	Parent node in Bayesian analysis
$P(t)$	The production rate
P_{BH}	Bottom hole pressure
P_P	Pore pressure
M	Sampling instance
m_i	Supremum of a function
n	Sample size
Nd	Nodes in Bayesian analysis that are not descendants
P	Number of estimable parameters
P*n*	*nth* percentile
Pr	Probability
r	Rank of a sample observation
R	Risk
R_c	Copula ratio
S	Number of simulations

S_σ	Strength in pounds per square inch
t	Time
T_x	Target value of process characteristic x
T_{max}	Maximum tolerable temperature
T_{min}	Minimum tolerable temperature
U	Uniform random variable
UF	Uncertainty factor
V	Uniform random variable
v	Shape parameter for PERT distribution
w	Shape parameter for PERT distribution
w_i	Akaike weights
$W(t, \tau_D)$	Deteriorating function
W_σ	Process variance deteriorating function
W_μ	Process mean deteriorating function
$x(t)$	Random variable denoting the process characteristic at time t
y	Process characteristic/variable

Greek letters

α	Loss function shape parameter
α_b	Shape parameter for Beta distribution
α_{DRAF}	Shape parameter for inverted Beta probability density function for DRAF
α_g	Scale parameter of Gamma distribution
α_{ins}	Fraction of insurance recovery
β_b	Shape parameter for Beta distribution
β_w	Weibull shape parameter
γ	Loss function shape parameter
$\Gamma(\cdot)$	Gamma function
γ_P	Parameter for PERT distribution
δ	Copula parameter

δ_μ	Shift in process characteristic mean
Δ	Shows difference between two values of a parameter. For example, distance from the target to the point where the maximum loss first occurs
Δ_i	Akaike differences
$\Delta\tau_C$	Process correction time period
$\Delta\tau_{dt}$	Process downtime (days)
$\Delta\tau_{LOC}$	Length of time between process deviation and loss of containment
$\Delta\tau_{rp}$	Recovery period (days)
$\Delta\tau_{SD}$	Length of time between process deviation and process shutdown
θ_w	Weibull scale parameter
$\theta_{i,k}$	Parameter used to show the performance of *i*th safety barrier when a failure level *k* occurs
η	Slope of the principal axis
θ_μ	Drift rate of the process characteristic mean
$\lambda_{\tau_C}^{-1}$	Exponential distribution mean for the correction time of process faults
λ_D^{-1}	Exponential distribution mean for the occurrence time of the assignable causes
λ_{LOC}^{-1}	Exponential distribution mean for the LOC time
λ_{SD}^{-1}	Exponential distribution mean for the shutdown time
ν_t	Degree of freedom for *t*-copula
μ_0	Initial process characteristic mean
$\mu_x(t)$	Mean value of process characteristic *x* at time *t*
μ_N	Mean of Normal distribution
μ_{LN}	Mean of lognormal distribution
π	Probability density function
ρ	Pearson correlation coefficient
ρ_s	Spearman's rank correlation coefficient
σ	Standard deviation

σ_N	Standard deviation for Normal distribution
Σ_{LN}	Standard deviation for Lognormal distribution
σ_0^2	Inherent (initial) variance of process characteristic x
$\sigma_x^2(t)$	Variance of process characteristic x at time t
τ	Kendall's rank correlation coefficient
τ_0	Process start time
τ_C	Correction time of process faults for an in-service process
τ_D	Occurrence time of the assignable causes and deviation of the process characteristic from specification limits to faulty state
τ_{LOC}	Process failure time due to loss of containment
τ_{SD}	Shutdown time

1. INTRODUCTION

1.1. Overview

Process safety is a disciplined framework for managing the integrity of operating systems and processes handling hazardous substances [1]. Unlike the occupational safety approach, which focuses on hazards that could result in health issues (e.g. slips, trips, and falls), process safety focuses on the prevention and mitigation of process hazards that may result in the release of chemicals or energy [2]. Such hazards could ultimately result in serious impacts including human health loss, environmental damage, asset loss and loss of production.

As industrialization progressed in the 20th century and more complex technologies developed over time, a series of intermittent catastrophes began to occure in different parts of the world [3]. However, it was mostly after highly publicized international disasters such as those in Flixborough (United Kingdom, 1974), Seveso (Italy, 1976), and Bhopal (India, 1984) [4] that governments and regulatory agencies began to establish what is now called Process Safety Management (PSM). Continued occurrence of major losses, with recent examples such as the Texas City refinery accident in 2005 [5] and the Gulf of Mexico's Deepwater Horizon oil spill in 2010 [6], has had a significant impact on the industry's approaches to modern process safety. Publications by Andrew Hopkins [7], Trevor Kletz [8], and Atherton & Gil [9] have served to raise process safety awareness and to publicize the lessons from these and other incidents [1].

The result of this increased awareness has been a global recognition of PSM as the primary approach for establishing the level of safety in operations required to manage high-hazard processes [1]. With this in mind, and also due to the evolution in regulatory thinking that integrated traditional occupational safety with process safety [10], several PSM standards and guidelines were developed by industry associations around the world. Some examples are the OSHA 29 CFR 1910.119 [11] and API RP 750 [12] in the United States, and the PSM Standard [13] published by CSChE in Canada. In Europe, since 1982, such approaches were integrated in the EU legislation, with the so-called "Seveso" Directives (Directive 82/501/EEC [14], Directive 96/82/EC [15], Directive 2012/18/EU [16]). All of these PSM programs cover the same basic requirements, although the number of program elements may vary depending on the criteria used [1].

Risk-based approaches were introduced to support the next generation of PSM programs. Some example of the projects conducted in Europe can be found in references [17–20], which are considered as the precursors of such approaches in chemical industries. In the United Kingdom and in The Netherlands, these methods are now required to support the implementation of Seveso Directives [2,21]. In the United States, risk-based approaches to PSM were introduced in 2007 by the Center for Chemical Process Safety (CCPS) [10]. The evolution of traditional PSM programs to risk-based approaches is described in the next section.

1.2. Risk-Based Process Safety

Although regulatory agencies around the globe have mandated a formal implementation of PSM, incident investigations continue to identify inadequate management system performance as a key contributor to incidents [22]. The disastrous Deepwater Horizon oil spill in the Gulf of Mexico in April 2010 has raised questions about the likelihood of such an event occurring again. There is a collective need for the chemical and major hazard sectors to demonstrate that risks are being adequately controlled, as the industry and regulators are often judged using the latest major incident to gain public attention [23]. Moreover, evolution of traditional PSM approaches is essential to avoid degradation of organizations' process safety performance.

Achieving process safety excellence requires identification of abnormal process situations and implementation of corrective actions before a serious incident occurs [10]. Built on analysis of the frequency and consequences of potential accidents, risk analysis is now an accepted tool by the oil and gas industry for evaluation of a PSM program performance [17,18]. "A risk-based approach reduces the potential for assigning an undue amount of resources to manage lower-risk events, thereby freeing up resources for tasks that address higher-risk events" [10]. Using risk-based process safety (RBPS), deficiencies of a PSM system can be identified and resources can be assigned accordingly to improve PSM practices. CCPS's RBPS guidelines [10] is perhaps the most important risk-based PSM program, which recognizes that all hazards and risks are not equal and that more resources should be focused on higher risks.

Several risk assessment techniques have been developed in recent years by different industries and regulatory agencies; some examples and reviews of conventional risk assessment can be found in [10,24,26,27]. Although conventional risk assessment methods have played an important role in identifying major risks and maintaining safety in process facilities, as disucussed in the following, they have the disadvantages of being static and using generic failure data [28].

Complex oil and gas facilities have hundreds of dynamic variables, deviations in which can affect the overall process risk [26]. Conditions in process facilities are dynamic, with changes in operating parameters often being reflected in changed operating procedures and equipment [29]. Feed variability, mechanical and operational integrity degradation, wrong settings, improper methods and human error can cause abnormal situations that can eventually result in severe consequences [30]. However, due to their static structures, conventional risk assessment approaches fail to capture the variation of risks as deviations or changes in the process and plant take place [28]. The following section identifies fundamental steps for conducting a dynamic risk assessment and reviews the major contributions in this area.

1.3. Dynamic Risk Assessment

Any process is subject to deterioration with time due to natural and assignable causes. A dynamic risk assessment (DRA) is defined in this work as a method that updates the estimated risk of a deteriorating process according to the new evidence from the performance of the control system, safety barriers, inspection and maintenance activities,

the human factor and procedures. Dynamic risk is usually measured as either a function of time or a function of one or more key process characteristics.

Figure 1.1 shows the flowchart of a typical DRA method. Almost all qualitative and quantitative risk analysis methods involve the first three steps identified in Figure 1.1 [29]. However, a DRA method includes an additional phase of monitoring and assessing abnormal process conditions to revise the estimated risk.

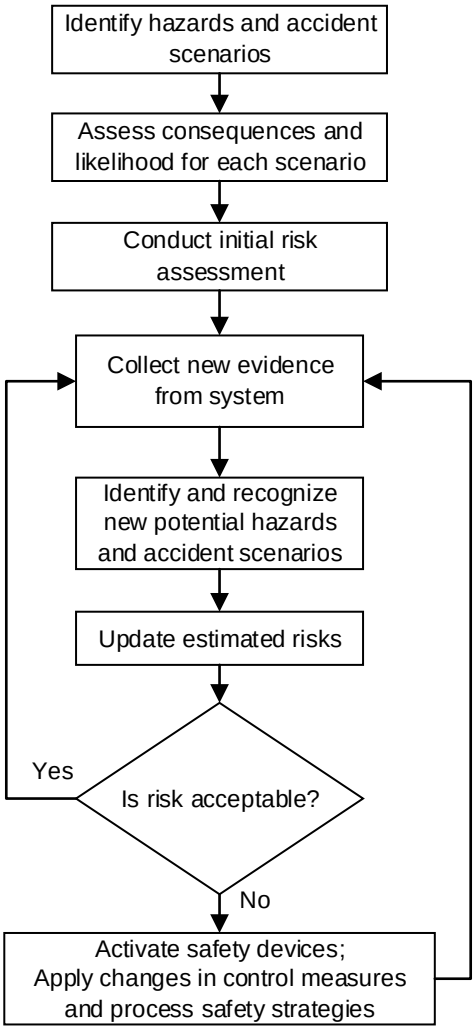


Figure 1.1. Typical dynamic risk analysis flowchart

There have been several contributions in recent years which propose and promote DRA methods [27,28,31–35]. A comprehensive review of these contributions shows that while all the presented approaches share the main steps identified in Figure 1.1, they can be distinguished based on the following three criteria:

i. Type of information used

ii. Risk updating mechanism

iii. Probability of failure (POF) updating vs. consequences of failure (COF) updating

Table 1.1 highlights the major contributions in dynamic risk analysis of process facilities and compares them based on the three criteria above. Table 1.1 also presents the advantages and disadvantages of each method. As can be seen from Table 1.1, Bayesian updating using process and failure histories and the bow-tie technique to incorporate newly identified potential hazards are two main approaches used in most DRA applications. However, due to the inherent limitations of each of the methods in Table 1.1, there still exist several knowledge and technological gaps in dynamic risk assessment. The current PhD research has been conducted to address these challenges, as will be discussed in the next section.

Table 1.1. Major contributions in dynamic risk analysis of process facilities

Method	Description	Related works
Bayesian	***Information used***: Process history, accident precursor data, alarm databases ***Updating Mechanism***: The Bayesian updating mechanism is used to update the prior beliefs about the accidents by incorporating new information from the system ***Advantages***: • Ability to handle uncertainty, multi-state variables, complex causal relationships, and sequentially dependent failures ***Disadvantages***: • High computational burden to construct conditional probability tables • Inability to model complex dependencies among variables • Application of deterministic and/or normally distributed probabilities	Meel and Seider [34,35] Pariyani et al. [32,36] Vinnem et al. [37] Kalantarnia et al. [28] Khakzad et al. [33]
Bow-tie	***Information used***: Accident precursor data; newly identified scenarios ***Updating Mechanism***: Continuous safety-related information retrieval is integrated with conventional bow-tie analysis to dynamically estimate the risk ***Advantages***: • Simple practical implementation ***Disadvantages***: • Limited ability to handle uncertainty, multi-state variables and dependent failures due to application of simple Boolean functions in bow-tie analysis • Application of deterministic probability values • Inability to model complex dependencies among variables	Paltrinieri et al. [38] Pasman and Rogers [39] CCPS [40] Khakzad et al. [41]
Principal component analysis (PCA)	***Information used***: Process history ***Updating Mechanism***: Probability of a fault is calculated based on the PCA filtered score, and the severity of the fault is a weighted average of the consequences of each variable in the score ***Advantages***: • Multivariate technique • Takes advantage of the correlation between process input and output • Extracts latent features from high dimensional data ***Disadvantages***: • Relies on linear models and assumes Gaussian noise • Unable to model complex dependencies among variables • Most applications require the process model	Jiang and Yang [42] Ge and Song [43] Zadakbar et al. [44]

Table 1.1. Major contributions in dynamic risk analysis of process facilities (*cont.*)

Method	Description	Related works
Risk barometer	***Information used***: Technical, operational and organizational indicators assessing deviation from optimal condition ***Updating mechanism***: Indicators are used to assess safety barrier performance on a regular basis, which in turn allows for continuous assessment of overall risk variation ***Advantages***: • Technical indicators are integrated with proactive operational/organizational indicators in order to assess early deviations potentially leading to unwanted events • Degrading safety barriers leading to critical risk increase may be identified and their improvement may be prioritized ***Disadvantages***: • Case-specific and partially based on expert judgment from operators • Relies on linear models • Based on relevant indicators whose collection may be irregular. Such irregularities may compromise the overall risk assessment	Paltrinieri et al. [31,45–47]
Loss functions	***Information used***: Deviation of key process characteristics from target values ***Updating Mechanism***: Loss functions are used to relate process deviations to economic losses ***Advantages***: • Provides a mechanism for real-time loss modelling • Promotes continuous improvement of process safety through proactive loss minimization ***Disadvantages***: • Selection of a proper loss function could be difficult for data-scarce processes	Hashemi et al. [30]† Zadakbar et al. [48] Ali [49] Pan and Chen [50]

† Reference [30] is based on the results of this thesis and is included in this table for completeness.

1.4. Motivations

The motivation of this thesis has been to bridge the main technological gaps between the existing methods and the requirements of an effective DRA method according to the definition provided in the previous section. Having evaluated the DRA methods in Table 1.1, these technological gaps are identified as follows:

i. The stochastic nature of losses is not taken into account. However, the uncertainty in loss model predictions can significantly affect the decision-making.

ii. The estimation of the operational loss of process facilities due to the deviation of key process characteristics (KPC) is usually ignored.

iii. The effect of reputational loss is ignored in almost all existing models and studies.

iv. Except for the loss function approach that can be used to revise estimated economical loss of a deviated process, all other methods may be considered dynamic only in estimating the probabilities of potential events.

v. In most approaches, it is assumed that a univariate key process characteristic can be assigned to a system.

vi. Deterministic point-based probability values are used in most applications, ignoring the uncertainty associated with probability estimations. In some recent developments in Bayesian Network (BN) and PCA-based approaches, normal distribution has been used as the marginal distribution. However, there is no doubt that the assumption of joint normality fails to yield suitable models in many applications.

vii. Either independent or linearly dependent variables are considered in all applications highlighted in Table 1.1. In other words, existing models fail to model complex non-linear dependencies among variables influencing system overall risk.

Although BN has been used as a general framework for analyzing causal influences among variables to analyze multivariate systems, the latter two challenges in the list above still exist. The inherent structural limitations of the BN and bow-tie approaches do not allow

consideration of multivariate systems with different marginal distributions with complex non-linear dependencies. From this discussion, it can be concluded that a combination of different mathematical tools would be required to enable incorporation of different types of process and failure information with different types of potentially complex and correlated marginal distributions.

This thesis uses an alternate and markedly different approach for constructing multivariate distributions using copulas to model operational losses using multivariate loss functions. The application of copula functions provides a flexible tool to capture stochastic dependency of complex systems by breaking down the dependency challenge into the estimation of marginal probability distributions and the estimation of dependency structure. Moreover, with copula modelling, the marginal distributions from different types can be combined, which is a significant improvement compared with alternative methods such as the use of multivariate distributions. However, copulas alone are unable to capture cause-effect relationships among random variables. Therefore, this research proposes the combination of BN analysis and copulas for dynamic probability estimation. The resulting Copula Bayesian Network (CBN) model provides an intuitively compelling framework for modeling causal relationships among (potentially) highly correlated variables with any level of dependence complexity.

The copula-based multivariate loss modelling proposed in Chapter 7 of this thesis and the CBN model for dynamic probability updating presented in Chapter 8 provide multivariate techniques that can dynamically update both the COF and POF of complex systems with

different types of discrete and continuous data and new information from the system, while addressing the technological gaps identified in the list above.

1.5. Scope and Objectives

The scope of this research covers both dynamic loss modelling and dynamic probabilistic analysis of process hazards in process facilities which may result in the release of chemicals or energy and loss of productivity. However, the development of a dynamic risk management framework is not within the scope of this thesis. The scope of the loss modelling parts of this thesis includes operational loss, reputational loss, and business interruption loss. This research also addresses the issue of loss aggregation to estimate overall loss. Detailed estimations of human health, environmental, and asset losses are not included in the scope of this work as they are extensively studied in the litarture. However, it is shown that the loss function approach and the proposed loss aggregation methodology are applicable for any type of loss. The models developed in this work are best suited for detailed quantitative risk assessment (QRA) of critical operations, such as offshore oil and gas development in harsh environments, where an accurate and precise risk estimation is required to ensure overall safety.

The proposed models perform the required analysis by answering these questions:

i. Is the system under control?
ii. What is the overall probability for an out-of-control state of the system?
iii. If the process is out-of-control, what will the consequences be?

iv. How to take the potential dependency among process and operational variables into consideration while estimating the process risk?

v. How should estimated loss and probability values be revised as new information from the system becomes available?

Having these research questions in mind, this work aims to fulfill the objectives identified in Figure 1.2 and described as follows:

i. Deveopment of dynamic risk-based methods for continuously improving the safety of operations.

ii. Improvement of existing loss modelling methods by considering losses as stochastic factors. Models are also provided to quantify operational loss, reputational loss and detailed business interruption loss, which are usually ignored in the existing loss modelling approaches.

iii. Determination of how process and operational deviations propagate through a deteriorating process and cause loss, where loss is defined as the discrepancy between the current situation and the ideal situation. The ultimate objective is to dynamically quantify the loss due to process deviations. Loss functions are used as a tool to achieve this objective.

iv. Development of a dynamic real-time probabilistic approach that can be used to estimate, revise and adapt the probability of operational losses of a deteriorating process according to the performance of the control system and safety barriers, considering complex non-linear dependencies among contributing factors. Integration of copula functions with the Bayesian Network is proposed to achieve

two main sub-objectives: (i) flexibility to assign different probability distributions to marginal risk factors; and (ii) incorporation of a potential complex, non-linear dependence structure among risk factors.

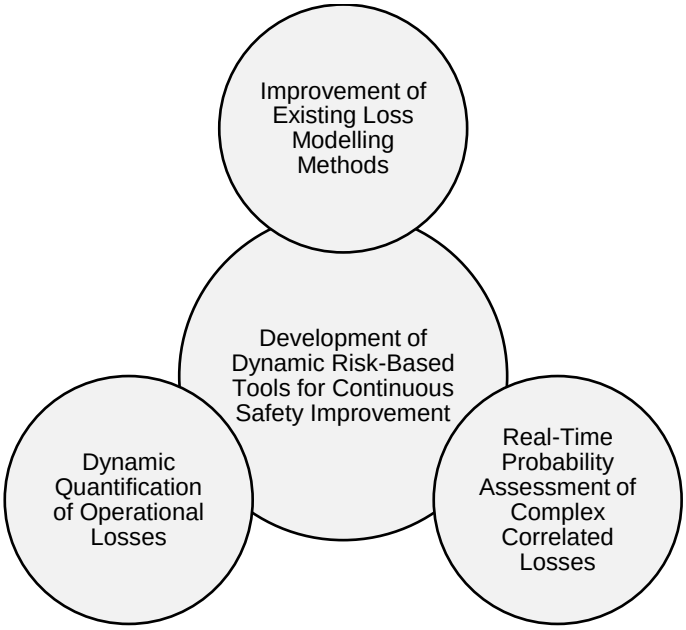


Figure 1.2. Research objectives

1.6. Contribution and Novelty

This section highlights the methodological and applicative contributions of this work and its significance by improving the existing methods in the area of dynamic risk assessment. One of the important methodological objectives of this research has been to comprehensively investigate risk assessment literature (not only in the area of the process industry, but also in other fields such as finance, economics, quality management, and actuarial literature) in order to identify potential methods and mathematical tools that can

be used to address identified challenges. The following is a brief description of the contributions and novelties of this research. The details of each identified contribution are provided in relevant chapters of the thesis.

1.6.1. Application of Loss Functions

The continued occurrence of major losses in the oil and gas industries highlights the fact that existing process safety management approaches are still far from what is necessary to avoid major losses. Therefore, there is a need for a renewed risk management approach for critical operations, focusing on the monitoring of process and operational deviations rather than reacting to accidents after they happen. Using the concept of loss function, this work provides a continuous improvement tool for complex operations by considering zero loss only for on-target operations to foster a zero-loss safety and quality culture. More details are provided in Chapters 2 and 3 of the thesis.

Moreover, using loss functions, this research provides an integrated risk-based safety and quality assessment and operational performance analysis tool for process operations; more details are provided in Chapter 3. Several static consequence analysis and loss modeling techniques for chemical processes have been proposed in the literature [24,51–53], none of which provide a dynamic consequence analysis approach. One of the main innovations of this research is the application of loss functions to monitor the real-time operational loss of process systems, which addresses the existing challenge in assigning monetary value to process deviations. More details are provided in Chapters 2, 3 and 7. Existing dynamic risk assessment models such as [28,32,35,54] use only the number of recorded incidents over

time to update risk. In addition to the observed incidents, the impact of process deviations as the root cause of the majority of process-related incidents is also incorporated in the proposed models in Chapters 2, 3 and 7 of the thesis by using the concept of loss functions.

1.6.2. Incorporating Complex Dependencies in Risk Assessment

Process safety and risk assessment are often multidimensional and hence require the joint modelling of several potentially correlated random variables. This research recognizes the importance of considering correlated variables used in risk assessment to avoid over-estimation or, in the worst case, underestimation of risk. A copula-based technique is used to model dependency among variables to improve the uncertainty analysis.

From modelling perspective, the main advantage of copulas compared with alternative methods, such as the use of multivariate distributions, is that the estimation of marginal distributions and the estimation of dependence structure can be performed separately. Moreover, with copula modelling, the marginal distributions from different families can be combined [55]. In practice, copula constructions often lead to a significant improvement in analyzing a system of correlated variables. Accordingly, there has been a growing interest in the application of copulas in the process industry, with its application mainly in the area of the risk analysis of safety systems [26,32,34–36].

The contributions of this research in terms of using copula functions are twofold. First, a new application of copula functions is provided to aggregate loss categories, considering the potential dependencies among them; details of this contribution are provided in Chapter 5. Second, methodologies are provided to estimate copula parameters and choose the best

copula for a specific application; details are provided in Chapter 7. The main objective of this part of the research is to present the successive steps required to use copulas for modelling the dependent losses and constructing multivariate distributions for specific purposes, including operational loss modelling.

1.6.3. Development of Multivariate Loss Modelling Techniques

Integrated operational loss modeling of process industries requires understanding the joint distribution of all key process characteristics and their correlations. Choosing and estimating a useful form for the marginal loss distribution of each variable in its domain is often a straightforward task. In contrast, as described in Chapter 7, other than the quadratic and inverted normal loss functions, univariate loss functions usually do not have a convenient multivariate generalization [55]. According to a review of the existing literature in the area of multivariate loss functions conducted by [56], it can be concluded that the existing research challenge is to develop a flexible framework to assign appropriate marginal loss functions to key process characteristics. The novelty of this research is to propose a methodology to construct the multivariate loss functions using copulas, which allows selection of any type of inverted probability loss function for the marginal losses, irrespective of their dependence structure. More details are provided in Chapter 7.

1.6.4. Development of Copula Bayesian Networks

Bayesian Network (BN) analysis has been used widely in process safety analysis for analyzing causal influences in multivariate systems and constructing joint distributions;

some recent examples are [36,57,58]. However, despite the broad scope of applicability, the following main shortcomings are identified for BN applications:

i. Deterministic point-based probability values are used in most BN applications, ignoring the uncertainty associated with probability estimations [33,59,60].
ii. Constructing the conditional probability tables (CPTs) to describe the strength of relationships quickly becomes very complex and difficult to compute as the number of parents and states increases [61].

This research proposes a model based on the Copula Bayesian Network (CBN) for multivariate safety analysis of process systems. The innovation of the proposed CBN model is in integrating the advantage of copula functions in modelling complex dependence structures with the cause-effect relationship reasoning of process variables using BNs. More discussion on the restrictions of traditional BN analysis and the advantages and novelty of the proposed CBN model are provided in Chapter 8.

1.7. Organization of the Thesis

This thesis is written in manuscript format (paper-based), including six journal papers and one peer-reviewed conference paper. Table 1.2 shows the papers published during the course of this research and the case study problem used to demonstrate the practical application of each part of the research.

Table 1.2. Organization of the thesis

Chapter Title	Supporting Paper Title	Case Study Problem
Chapter 1: Introduction	Not applicable (NA)	NA
Chapter 2: Loss Functions	Loss functions and their applications in process safety assessment. *Process Safety Progress*, 33(3), 285–291 (2014)	Reactor process (early warning design)
Chapter 3: Applications of Loss Functions in Risk Assessment	Risk-based operational performance analysis using loss functions. *Chemical Engineering Science*, 116, 99–108 (2014)	Continuous Stirred Tank Reactor (CSTR)
Chapter 4: Dependency in Multivariate Process Risk Assessment	Correlation and dependency in multivariate process risk assessment. In *9th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes*, Paris (2015)	Distillation process
Chapter 5: Loss Aggregation	Loss scenario analysis and loss aggregation for process facilities. *Chemical Engineering Science*, 128(May 2015), 119–129 (2015)	Distillation column
Chapter 6: Business Loss Modelling	Probabilistic modelling of business interruption and reputational losses for process facilities. *Process Safety Progress*, 34(4), 373–382 (2015)	Distillation column
Chapter 7: Multivariate Loss Functions	Operational loss modelling for process facilities using multivariate loss functions. *Chemical Engineering Research and Design*, 104, 333–345 (2015)	Simulation study; De-ethanizer column
Chapter 8: Multivariate Probabilistic Safety Analysis	Multivariate probabilistic safety analysis of process facilities using the Copula Bayesian Network model. *Computers and Chemical Engineering*, 93 (2016) 128–142	Offshore managed pressure drilling (MPD)
Chapter 9: Summary, Conclusions and Recommendations	NA	NA

The outline of each chapter is explained below:

Chapter 2 reviews the potential applications of loss functions in process industries. Loss functions are used to define the relationship between process deviations and system loss. As an example application, consequence assessment using loss functions is incorporated into a risk-based warning system design model to analyze warnings associated with process deviations.

Chapter 3 proposes a risk-based process performance assessment methodology using loss functions. The demand rate adjustment factor is used to model the effect of process deviations on the failure probability of safety systems. This chapter highlights the use of the loss function approach to continuously update the system loss based on the current value of the characteristic variables.

Chapter 4 discusses the problems with correlated variables used in risk assessment and presents a copula-based technique to model dependency among variables and improve the uncertainty analysis. Using the copula approach, capturing the dependence structure among different risk factors and estimating the univariate risk marginals can be separated.

Chapter 5 presents an application of the copula functions, and their integration with the Monte Carlo (MC) approach, to address the existing challenge of loss aggregation for multiple-loss scenarios. The proposed loss aggregation provides a flexible and realistic approach to constructing a joint multivariate distribution of the losses by considering their interdependence.

Chapter 6 presents probabilistic models to estimate business losses due to abnormal situations in process facilities. The main elements of business loss are identified as business interruption loss and reputational loss and models are presented for each of these two elements. Copula functions are then used to develop the distribution of the aggregate loss, considering the correlation between business interruption and reputational losses.

Chapter 7 presents another application of copula functions to develop multivariate loss functions in order to measure the operational loss of process facilities. Methods are also presented to address two main challenges of the application of copula approaches, which

are parameter estimation and copula selection. The maximum likelihood evaluation method is used to estimate the copula parameters. Akaike's information criterion (AIC) is then applied to rank the copula models and choose the best fitting copula.

Chapter 8 presents a novel model based on the Copula Bayesian Network (CBN) for multivariate safety analysis of process systems. This offers a great flexibility in probabilistic analysis of individual risk factors while considering their uncertainty and stochastic dependence. Chapter 9 reports the summary of the thesis and the main conclusions drawn through this work. Recommendations for future work are presented at the end of Chapter 9.

1.9. References

[1] CAPP. Process safety management. Calgary: The Canadian Association of Petroleum Producers (CAPP); 2014. doi:10.1002/prs.11678.

[2] HSE. The Control of Major Accident Hazards (COMAH) Regulations. Third. London: Health and Safety Executive (HSE); 2015.

[3] Macza M. A Canadian Perspective of the History of Process Safety Management Legislation 2008:1–22.

[4] Marsh Risk Consulting. The 100 Largest Losses. London: 2014.

[5] CSB. Investigation Report for BP Texas City Refinery Explosion. Washington, DC: The U.S. Chemical Safety Board; 2007.

[6] CSB. Investigation Report Overview: Explosion and Fire at the Macondo Well. Washington, DC: The U.S. Chemical Safety Board; 2014.

[7] Hopkins A. Lessons from Longford. Sydney, Australia: CCH Australia; 2000.

[8] Kletz T. Still Going Wrong! : Case Histories of Process Plant Disasters and How They Could Have Been Avoided. Burlington, MA: Gulf Professional Publishing; 2004. doi:10.1016/B978-075067709-7/50058-5.

[9] Atherton J, Gil F. Incidents That Define Process Safety. New York: Center for Chemical Process Safety (CCPS); 2008.

[10] CCPS. Guidelines for Risk Based Process Safety. Hoboken, New Jersey: Center for Chemical Process Safety. John Wiley & Sons, Inc.; 2007.

[11] OSHA. Process Safety Management of Highly Hazardous Chemicals (29 CFR 1910.119). U.S.

Occupational Safety and Health Administration; 1992.

[12] API. RP 750: Management of Process Hazards. Washington: American Petroleum Institute; 1990.

[13] CSChE. Process Safety Management Standard. First. Ottawa, ON: Canadian Society for Chemical Engineering; 2012.

[14] EC. Council Directive 82/501/EEC of 24 June 1982 on the major-accident hazardsof certain industrial activities. Brussels: Official Journal of the European Communities L230/25; 1982.

[15] EC. Council Directive 96/82/EC of 9 December 1996 on the control of major-accident hazards involving dangerous substances. Brussels: Official Journal of the European Communities L10/13; 1997.

[16] EC. European Parliament and Council Directive 2012/18/EU of 4 July 2012 on control of major-accident hazards involving dangerous substances, amending and subsequently repealing. Brussels: Official Journal of the European Communities, L 197/1; 2012.

[17] HSE. Canvey: A Second Report. A Review of the Potential Hazards from Operations in the Canvey Island/Thurrock Area Three Years after Publication of the Canvey Report. HM Stationery Office, London: Health and Safety Executive; 1981.

[18] HSE. Canvey: An Investigation of Potential Hazards from Operations in the Canvey Island/Thurrock Area. HM Stationery Office, London: Health and Safety Executive (HSE); 1978.

[19] Rijnmond Public Authority. Risk Analysis of Six Potentially Hazardous Industrial Objects in the Rijnmond Area, a Pilot Study. Reidel, Dordrecht (NL): Springer Science & Business Media; 1982.

[20] Egidi D, Foraboschi FP, Spadoni G, Amendola A. The ARIPAR project: analysis of the major accident risks connected with industrial and transportation activities in the Ravenna area. Reliab Eng Syst Saf 1995;49:75–89. doi:10.1016/0951-8320(95)00026-X.

[21] Uijt de Haag PAM, Ale BJM. Guidelines for Quantitative Risk Assessment (Purple Book). The Hague (NL): Committee for the Prevention of Disasters; 1999.

[22] CCPS. Risked Based Process Safety Overview. New York: American Institute of Chemical Engineers (AIChE); 2014.

[23] HSE. Developing process safety indicators. London: Health and Safety Executive; 2006.

[24] API. Recommended Practice 581: Risk-Based Inspection Technology. 2nd ed. Washington: American Petroleum Institute; 2008.

[25] CCPS. Guidelines for Chemical Process Quantitative Risk Analysis. Second. New York: Center for Chemical Process Safety of the American Institute of Chemical Engineers; 2000.

[26] Oktem UG, Seider WD, Soroush M, Pariyani A. Improve Process Safety with Near-Miss Analysis. Chem Eng Prog 2013;May:20–7.

[27] Khan F, Rathnayaka S, Ahmed S. Methods and models in process safety and risk management: Past, present and future. Process Saf Environ Prot 2015;98:116–47.

[28] Kalantarnia M, Khan F, Hawboldt K. Dynamic risk assessment using failure assessment and Bayesian theory. J Loss Prev Process Ind 2009;22:600–6. doi:10.1016/j.jlp.2009.04.006.

[29] NOPSEMA. Guidance note N-04300-GN0165 Revision 4 - Risk Assessment. Perth, Australia: National Offshore Petroleum Safety and Environmental Management Authority; 2012.

[30] Hashemi SJ, Ahmed S, Khan F. Risk-based operational performance analysis using loss functions. Chem Eng Sci 2014;116:99–108. doi:10.1016/j.ces.2014.04.042.

[31] Paltrinieri N, Scarponi G. Addressing Dynamic Risk in the Petroleum Industry by Means of

Innovative Analysis Solutions. Chem Eng Trans 2014;36:451–6. doi:10.3303/CET1436076.

[32] Pariyani A, Seider WD, Oktem UG, Soroush M. Dynamic Risk Analysis Using Alarm Databases to Improve Process Safety and Product Quality: Part I - Data Compaction. AIChE J 2012;58:812–25. doi:10.1002/aic.12643.

[33] Khakzad N, Khan F, Amyotte P. Dynamic safety analysis of process systems by mapping bow-tie into Bayesian network. Process Saf Environ Prot 2013;91:46–53. doi:10.1016/j.psep.2012.01.005.

[34] Meel A, Seider WD. Plant-specific dynamic failure assessment using Bayesian theory. Chem Eng Sci 2006;61:7036–56. doi:10.1016/j.ces.2006.07.007.

[35] Meel A, Seider WD. Real-time risk analysis of safety systems. Comput Chem Eng 2008;32:827–40. doi:10.1016/j.compchemeng.2007.03.006.

[36] Pariyani A, Seider WD, Oktem UG, Soroush M. Dynamic Risk Analysis Using Alarm Databases to Improve Process Safety and Product Quality: Part II - Bayesian Analysis. AIChE J 2012;58:826–41. doi:10.1002/aic.12642.

[37] Vinnem JE, Bye R, Gran BA, Kongsvik T, Nyheim OM, Okstad EH, et al. Risk modelling of maintenance work on major process equipment on offshore petroleum installations. J Loss Prev Process Ind 2012;25:274–92. doi:10.1016/j.jlp.2011.11.001.

[38] Paltrinieri N, Tugnoli A, Buston J, Wardman M. DyPASI Methodology : from Information Retrieval to Integration of HAZID Process. Chem Eng Trans 2013;32:433–8. doi:10.3303/CET1332073.

[39] Pasman H, Rogers W. How can we use the information provided by process safety performance indicators? Possibilities and limitations. J Loss Prev Process Ind 2014;30:197–206. doi:10.1016/j.jlp.2013.06.001.

[40] CCPS. Guidelines for Hazard Evaluation Procedures. New York: American Institute of Chemical Engineers (AIChE); 2008.

[41] Khakzad N, Khan F, Amyotte P. Dynamic risk analysis using bow-tie approach. Reliab Eng Syst Saf 2012;104:36–44. doi:10.1016/j.ress.2012.04.003.

[42] Jiang Q, Yan X. Nonlinear plant-wide process monitoring using MI-spectral clustering and Bayesian inference-based multiblock KPCA. J Process Control 2015;32:38–50. doi:10.1016/j.jprocont.2015.04.014.

[43] Ge Z, Song Z. Distributed PCA model for plant-wide process monitoring. Ind Eng Chem Res 2013;52:1947–57. doi:10.1021/ie301945s.

[44] Zadakbar O, Imtiaz S, Khan F. Dynamic Risk Assessment and Fault Detection Using Principal Component Analysis. Ind Eng Chem Res 2012;52:809–16. doi:10.1021/ie202880w.

[45] Paltrinieri N, Hauge S, Dionisio M, Nelson WR. Towards a dynamic risk and barrier assessment in an IO context. Safety, Reliab. Risk Anal. Beyond Horiz. - Proc. Eur. Saf. Reliab. Conf. ESREL 2013, Amsterdam, Netherlands: 2014, p. 1915–23. doi:10.1201/b15938-293.

[46] Paltrinieri N, Hauge S, Nelson WR. Dynamic barrier management : A case of sand erosion integrity. Saf. Reliab. Complex Eng. Syst. Proc. Eur. Saf. Reliab. Conf. ESREL 2015, Zurich, Switzerland: 2015, p. 523–31.

[47] Paltrinieri N, Hokstad P. Dynamic risk assessment: Development of a basic structure. Saf. Reliab. Methodol. Appl. - Proc. Eur. Saf. Reliab. Conf. ESREL 2014, Wroclaw, Poland: 2015, p. 1385–92. doi:10.1201/b17399-191.

[48] Zadakbar O, Khan F, Imtiaz S. Dynamic Risk Assessment of a Nonlinear Non-Gaussian System Using a Particle Filter and Detailed Consequence Analysis. Can J Chem Eng 2015;93:1201–11. doi:10.1002/cjce.22212.

[49] Ali S. Mixture of the inverse Rayleigh distribution: Properties and estimation in a Bayesian framework. Appl Math Model 2015;39:515–30. doi:10.1016/j.apm.2014.05.039.

[50] Pan J-N, Chen S-C. A loss-function based approach for evaluating reliability improvement of an engineering design. Expert Syst Appl 2013;40:5703–8. doi:10.1016/j.eswa.2013.04.032.

[51] CCPS. Guidelines for Consequence Analysis of Chemical Releases. New York: Center for Chemical Process Safety of the American Institute of Chemical Engineers; 1999.

[52] Dadashzadeh M, Khan F, Hawboldt K, Amyotte P. An integrated approach for fire and explosion consequence modelling. Fire Saf J 2013.

[53] Arunraj NS, Maiti J. A methodology for overall consequence modeling in chemical industry. J Hazard Mater 2009;169:556–74. doi:10.1016/j.jhazmat.2009.03.133.

[54] Khakzad N, Khan F, Amyotte P. Dynamic risk analysis using bow-tie approach. Reliab Eng Syst Saf 2012;104:36–44. doi:10.1016/j.ress.2012.04.003.

[55] Elidan G. Copula Bayesian Networks. Adv. neural Inf. Process. Syst. (NIPS 2010), Vancouver: 2010, p. 559–67.

[56] Hashemi SJ, Ahmed S, Khan F. Loss functions and their applications in process safety assessment. Process Saf Prog 2014;33:285–91. doi:10.1002/prs.11659.

[57] Bauer AX. Pair-copula constructions for non-Gaussian Bayesian networks (Doctoral dissertation). Technische Universitat Munchen, 2013.

[58] Khakzad N, Khan F, Amyotte P, Cozzani V. Domino effect analysis using Bayesian networks. Risk Anal 2013;33:292–306. doi:10.1111/j.1539-6924.2012.01854.x.

[59] Yuan Z, Khakzad N, Khan F, Amyotte P. Risk Analysis of Dust Explosion Scenarios Using Bayesian Networks. Risk Anal 2015;35:278–91. doi:10.1111/risa.12283.

[60] Abimbola M, Khan F, Khakzad N. Risk-based safety analysis of well integrity operations. Saf Sci 2016;84:149–60. doi:10.1016/j.ssci.2015.12.009.

[61] Mohseni Ahooyi T, Arbogast JE, Soroush M. Applications of the Rolling Pin Method. 1. An Efficient Alternative to Bayesian Network Modeling and Inference. Ind Eng Chem Res 2014;54:4316–25. doi:10.1021/ie503585m.

[62] Kraus RS. Developing a Process Safety Management Programme. Int Labour Organ 2011. http://www.iloencyclopaedia.org/component/k2/127-77-chemical-processing/developing-a-process-safety-management-programme (accessed December 22, 2015).

2. LOSS FUNCTIONS AND THEIR APPLICATIONS IN PROCESS SAFETY ASSESSMENT[1]

Preface

A version of this manuscript has been published in the Journal of Process Safety Progress. I am the primary author of this paper. Along with the co-authors, Faisal Khan and Salim Ahmed, I developed the conceptual model. I carried out most of the literature review, data collection and the comparison of loss functions. I prepared the first draft of the manuscript and subsequently revised the manuscript based on the co-authors' feedback and also the peer review process. The co-author Faisal Khan helped in developing and testing the concepts/models, reviewed and corrected the models and results, and contributed in preparing, reviewing and revising the manuscript. The co-author Salim Ahmed contributed through support in the development, testing and improvement of the models. Salim Ahmed also assisted in reviewing and revising the manuscript.

Abstract

Process deviations, along with failure of control systems and protection layers, result in safety and quality loss in plant operations. This paper proposes an operational risk-based warning system design methodology based on overall system loss. Loss functions are used to define the relationship between process deviations and system loss. For this purpose,

[1] Hashemi et al. Process Safety Progress 2014;33:285–91.

properties associated with quadratic loss function and a set of inverted probability loss functions are investigated and compared. The results suggest that loss functions can be used in a novel way to assess operational stability and system safety. The proposed consequence assessment methodology using loss functions is then incorporated into a risk-based warning system design model to analyze warnings associated with process deviations. A simulated case study is presented to demonstrate potential application of the proposed methodology; the study examines the response to a temperature surge for a reactor system.

Keywords: process safety management; alarm warning; alarm management; risk assessment

2.1. Introduction

An important aspect of any industrial operation is conformance to standards. It relates to how closely the operational performance and process safety, as well as quality of the operation and final products, match the design specifications. In particular to the process industry, meeting the target performance is of great importance due to increasing demands for higher efficiency and strict environmental regulations. Achieving conformance requires investment on appraisal, prevention, and improvement techniques. Evaluating the loss incurred due to deviation from the expected level of conformance provides a new process performance indicator to justify both the cost of investment on continuous improvement of operations and a renewed commitment to run a safer and greener process.

Taguchi [1] proposed a quadratic loss function to illustrate losses to society associated with deviations of quality characteristics from their operational targets in industrial applications. Taguchi's quality philosophy states that there is a cost for any finite deviation. Commonly referred to as Taguchi's loss function, this loss imposes maintenance and repair costs to consumers, warranty and scrap costs to manufacturers, and pollution and environmental costs to the society [2].

The loss function (LF) approach is widely used to quantify losses associated with deviation from target value for various purposes such as economic design of specification region [3], obtaining optimal inventory and investment policies [4], business decision-making [5], quality assurance [6], risk assessment [7], marine and offshore safety [8], supplier selection [9] and reliability settings [10]. There are many other reported applications of LFs to quantify unobservable variability costs. These are mostly in the area of quality engineering; interested readers are referred to [4,7,9,11,12].

Despite successful application of the LFs, an obvious difficulty regarding their use is to determine their exact form. Various LFs have been discussed in the literature and in some applications different types of LFs are used for similar purposes. While many researchers agree in principle with the premise of using LFs in the context of managing continuous improvement, there is no general agreement on the specific form of the LF to be used for this purpose [13]. As a result, one of the most critical issues encountered in any application of LFs is the selection of a proper LF to relate a key characteristic of a system to its performance. Using inappropriate LFs may lead to inaccurate results that either underestimate or overestimate the loss [14].

The fundamental objective of this work is to study the advantageous application of LFs in process safety management. In this paper, the general basis of LFs and a review of relevant literature on LFs are discussed in Section 2.2. Use of LFs in different industrial applications is analyzed in Section 2.3. The potential applications in the context of operational risk-based process safety management and associated challenges are discussed in Section 2.4. A methodology for warning system design by integrating the loss functions in risk determination is proposed in Section 2.5. Numerical comparisons of the LFs by examining the response to a temperature surge for a reactor system and a sensitivity analysis to study the impact of changing the shape of loss functions on the results are provided in Section 2.6. Conclusions and recommendations are drawn in Section 2.7.

2.2. Loss Functions

2.2.1. Univariate Loss Functions

Taguchi [1] defined the quadratic loss function (QLF) as $L(y) = B(y-T)^2$ where y denotes the quality characteristic, $L(y)$ is the actual loss at y, T is the target value, and B is a constant. With its infinite maximum loss and symmetric shape, QLF is inadequate for some applications in describing the loss [15]. Joseph [16] proposed a set of LFs based on Taguchi's loss concept to address problems requiring asymmetric shapes . In practice, the maximum loss is generally finite. Truncating the QLF at the points where the function intersects the maximum loss addresses this concern [6]. To implement this idea, the concept of inverted normal loss function (INLF) was proposed by Spiring [6] which has a bounded value from above and its supremum can be specified by the user. Sun et al. [17] refined the

INLF further and developed the modified INLF (MINLF). This LF has a shape parameter γ that is specified by the user and its value determines the slope of the function in the neighborhood of the target value. Later, Spiring and Leung [5] extended the concept of INLF further to other inverted probability density functions including inverted beta loss functions (IBLF) which not only provide the traditional properties of LFs but also include the asymmetrical loss cases [5,11,18].

Table 2.1 provides the formulations for the most widely used univariate LFs in the literature. In Table 2.1, y denotes the quality characteristic, $L(y)$ is the actual loss at y, T is the target value, EML is the estimated maximum loss, Δ is the distance from the target to the point where the maximum loss EML first occurs, and α and γ are shape parameters and need to be determined from additional information for MINLF and IBLF.

Table 2.1. Listing of univariate loss function formulations

Type of LF	Reference	Formulation of Loss Function
QLF	[1]	$L(y) = B(y-T)^2$ where $B = EML/\Delta^2$
INLF	[6]	$L(y) = EML\{1-exp(-(y-T)^2/2\gamma^2)\}$ where $\gamma = \Delta/4$
MINLF	[17]	$L(y) = \frac{EML_{\Delta}}{1-exp\{-0.5(\Delta/\gamma)^2\}}\{1-exp(-(y-T)^2/2\gamma^2)\}$
IBLF	[11]	$L(y) = EML\{1-C_L[y(1-y)^{(1-T)/T}]^{(\alpha-1)}\}$ where $C_L = [T(1-T)^{1-T/T}]^{1-\alpha}$

2.2.2. Multivariate Loss Functions

While the existing literature widely covers the univariate LFs, there are some attempts, mainly in the area of quality engineering, to extend the LF concept to multivariate problems. A review of some multivariate LFs can be found in [19–21]. Traditionally, assigning a weight for each response has been considered to be an effective way to a

multivariate loss function [20]. The main drawback of these approaches is that in real-life industrial applications, quality characteristics are often not independent or additive [19].

Pignatiello [22] and Artiles-Leon [23] are among the early researchers reporting the application of multivariate LFs and their works are the bases for many later research [21]. Pignatiello [22] presented a quadratic LF for multi-response problems and established a predictive regression model by using controllable variables. As argued in [20], in Pignatiello's approach, it is difficult to determine the cost matrix and additional experimental observations may be required. Artiles-Leon [23] included specification limits in the LF itself. Later, Ma [24] improved the Artiles-Leon's work by applying principal component analysis (PCA) to consider correlation structures among the various responses. Hsu [20] and Su and Tong [25] have also used PCA to use Taguchi's method for multiple system characteristics.

Spiring [6] extended the univariate INLF to the case where two characteristics are of interest in assessing loss. However, the dependency between the parameters is not considered in [6]. Extensions of the concept of QLF for considering three main categories of multiple quality characteristics including nominal-the-best, smaller-the-better, and larger-the-better cases are reported in [14,19,26] where it is assumed that the quality characteristics follow multivariate distributions. To consider the dependency among quality characteristics, the Bessel function and a correlation coefficient were used to obtain the joint probability distribution of multiple quality characteristics when calculating the expected loss.

2.3. Use of Loss Functions

In the past two decades the use of loss functions for measuring quality costs for the purpose of process quality improvement has rekindled interest for researchers and practitioners alike [4]. One of the active areas of research has been the application of LFs for process improvement and production planning. For a review of the literature on application of LFs in production planning, economic lot-sizing, and inspection planning interested readers are referred to [4] and [27].

LFs have also been used to optimize the expected loss to the customer as well as to the manufacturer. Abdul-Kader [27] presented a review of the literature on product and process optimization and loss minimization models and proposed an integrated cost model composed of a tolerance model and an investment model using Taguchi's QLF. However, only the single variable optimization problems are considered in [27]. Raiman [28] emphasized the importance of the multivariate quality characteristic case and concluded that the total loss for a given process can be obtained by adding the losses resulting from each quality characteristic with each loss based on an independent quality characteristic and its associated loss. References [3,15,29,30] have also studied and proposed process optimization methods and loss minimization models for multiple quality characteristics problems. To relax the assumption of independency, references [20] and [25] presented an approach based on PCA to optimize the multi-response problems.

Pan [31] proposed a LF-based risk assessment method by linking the process capability indices and LFs. In [31] the univariate capability indices of processes and LFs describe the likelihood and consequences of non-conformities in the system, respectively. More

recently, Pan and Chen [7] extended the methodology in [31] to multivariate cases and developed a correlated risk assessment technique for manufacturing and environmental systems. This approach is based on Pignatiello's quadratic multivariate LF, where in addition to restrictions of QLF in modeling loss, it requires to determine the cost matrix which could be difficult to obtain. However, this work is among the early efforts to apply LFs in quantitative risk assessment.

2.4. Application of Loss Functions in Process Safety Assessment

The work by Sii et al. [8] to apply the Taguchi concept in maritime safety engineering systems is a prime effort to use LF for improvement of system safety. More recently, a work by Pan and Chen [7] to find the relationship between process capability indices and LFs to develop a LF based risk assessment approach has been reported. However, there appears to be virtually no study that uses LFs for risk-based safety assessment and management for process industries.

The limited application of LFs in process systems can be related, in part, to the original intention of their application to optimize engineering processes [1]. Compared with manufacturing industries, deviation of the key process characteristic from the target markedly increases the probability of process safety loss. This is mainly due to hazardous operations under high pressure and temperature in process industries and involvement of dangerous substances. In addition to the original intention of LFs to optimize manufacturing settings, when comparing process industries with its manufacturing counterparts, there are some reasons why the LF approach have not been employed in

process safety studies. The main reasons include the difficulty in measuring the safety of a system precisely and relating it to monetary values. In addition, process systems comprise of more unknown, or imprecisely known, parameters [8].

In traditional process safety assessment, no loss with respect to product quality and system safety is assumed until the key process characteristic deviates from the operational boundaries. However, in the LF approach a process imparts "no loss" only if the process characteristic of interest equals its target. In terms of system safety, as the deviation from the target increases there is a larger possible demand for activation and functionality of control and safety systems. Due to inevitable uncertainty in the successful operation of control and safety systems, the deviation of the process characteristic increases the failure possibility of safety systems. Application of the LF approach in process safety assessment and consideration of zero loss only for on-target operation can lead to the direction of continuous safety improvement of process operations.

Process safety management is widely recognized and credited for accident risk reduction and improved operational performance. However, process safety management appear to have stagnated within many organizations due to application of inadequate management system performance indicators such as incident investigation [32]. Key to achieving a safe and economic process operation is the application of leading indicators that focus on the process safety and quality elements that matter, thereby providing a true measure of how an asset is performing.

With the purpose of developing a process safety performance leading indicator, a LF approach is proposed in this work to quantify loss due to deviation of process

characteristics from their target operation. Current loss modelling and consequence analysis methodologies are focused only on obvious losses due to production downtime and release of material and energy. Using a loss function approach, other aspects of system loss such as the loss of product quality due to process deviations and their potential impacts on increasing the chance of experiencing more severe safety related consequences can be quantified. The measured loss is then integrated in a risk estimation model which provides a leading indicator of process safety performance based on the current process state. The estimated risk is finally used as a criterion to annunciate, prioritize, and analyze warnings which are generated as a result of process deviations.

2.5. Risk-Based Warning System Design

Integration of loss function based consequence analysis into an operational risk assessment model is proposed to be used in alarm system design. An alarm system is one of the critical layers of protection in a process safety system. Alarms prevent the escalation of abnormal situations and minimize the demand for activation of safety instrumented systems (SIS) such as a trip or an emergency shutdown device (ESD) [33]. However, the appearance of too many alarms, also known as alarm flooding, when the process approaches an abnormal situation has been reported as a contributor to abnormal events [34]. The discussion on the shortcoming of the current alarm management systems can be found in [35,36].

To address the alarm flooding problem in process facilities, Chang et al. [33] proposed a quantitative risk-based alarm design approach for annunciation and prioritization of alarms. An integrated model consisting of the probability and the impact of the potential

hazards, and the process safety time is used to estimate the risk according to Equation 2.1 [33].

$$Risk = Pr \times L \times 100^{\left(1-t/60\right)} \tag{2.1}$$

where *Risk* is the final risk of an alert, *Pr* is the probability of the potential hazard, *L* is the severity or impact of the consequences, and *t* is the process safety time between the initiating event and the occurrence of a hazardous event in the unit of minutes [33,34]. Scaling factors based on expert's judgment and process knowledge to determine the $Pr \times L$ score for each hazard is used in [34]. Although this approach to assign $Pr \times L$ score has the advantage of simplicity and consistency when analyzing multiple hazards, it adds uncertainty to the calculated risk.

We propose to integrate the loss function approach to estimate system loss as a function of process deviation into risk equation. The formulation of the proposed model is in accordance with Equation 2.2:

$$Risk = Pr \times L\left(x,T\right) \times 100^{\left(1-t/60\right)} \tag{2.2}$$

where *x* is the key process characteristic, *T* is the target operating value of the process characteristic of interest, *Pr* is the probability of the potential hazard, *t* is the process safety time, and $L(x,T)$ is the overall loss calculated by using loss functions (listed in Table 2.1). The method continuously evaluates the risk associated with the current state of a process and issues warnings based on the estimated risk. By comparing the estimated risk with a predefined threshold risk, warnings are issued and categorized into alerts and alarms which will be annunciated on different displays. In addition to affecting the system loss, process

deviations affect the probabilities of failure of safety systems which is the subject for another ongoing research.

The proposed approach determines risk in dollar value instead of a risk ranking based on the $Pr \times L$ score, thereby providing a more sensible criterion to make operational risk-based decisions. Choosing the threshold risk can be based on the tolerable risk limit of the system which is a challenging task in industrial applications. Wrong assignment of the threshold risk may have adverse effect on process safety. It may result in too many alarms or, on the other hand, may downgrade an alarm to an alert. Another advantage of the proposed method in representing the risk in dollar values is to provide a convenient tool for business managers to assign a threshold risk for a particular process system.

Despite several advantages of using the LF approach in risk-based warning system design, as discussed above, the choice of the LF could be a practical difficulty. The effect of changing the shape of the LFs on the estimated risk and a comparison of different LFs are discussed in the next section. The procedure is demonstrated by application of the methodology to a reactor system to examine the response to a temperature surge. This comparison will help industrial practitioners in choosing the appropriate LF for their applications.

2.6. Application and Analysis of Results

2.6.1. Application of Loss Functions to a Reactor System

To compare the performance of different loss functions, namely QLF, INLF, MINLF, and IBLF (see Table 2.1), system loss associated with temperature deviations in a reactor

system is calculated using the above mentioned LFs. This example is taken from [33]. The first step is to identify the potential hazards where the hazards related to the temperature surge in the reactor are identified as overheating and under-heating of the reactor liquid. The nomenclature of different parameters used in the calculations and the user-defined values are listed in Table 2.2.

Table 2.2. Reactor process information used to develop LFs

Symbol	Description	Value
T_T	Target reaction temperature of the reactor	120 ^oC
T_{HA}	Temperature set point for high temperature alarm	130 ^oC
USL	Upper specified limit for liquid temperature associated with high-high temperature alarm (set point for ESD system)	150 ^oC
LSL	Lower specified limit for liquid temperature associated with low temperature alarm	100 ^oC
T_{max}	Maximum tolerable liquid temperature; it is assumed that at this temperature the reactor fails catastrophically	200 ^oC
T_{min}	Minimum tolerable liquid temperature; it is assumed that at this temperature ESD system will be activated	80 ^oC
EML_{USL}	Loss associated with liquid temperature at USL	2,000,000 US$
EML_{LSL}	Loss associated with liquid temperature at LSL	1,000,000 US$
EML_1	Estimated maximum loss due to high temperature scenario in the reactor including the production, asset, environment clean-up, and human health losses.	10,000,000 US$
EML_2	Estimated maximum loss due to low temperature scenario in the reactor due to process downtime	2,000,000 US$

As the reactor temperature deviates more from the desired target reaction temperature (120^oC) and upon failure of the control system and other protection layers, the system experiences more loss. The identification and estimation of the applicable losses are based on the generated accident scenarios associated with the potential hazard. Two general types of loss are expected due to the deviation of the reactor temperature:

1. Quality loss: the loss due to deviation of the temperature from the target due to the occurrence of process near misses. It is assumed that the deviation of the reactor

temperature within the safe operating window only causes the loss of quality of final product.

2. Accident loss: loss due to occurrence of process incidents and accidents ranging from the production loss due to process shutdown to a combination of production, asset, injury and fatality, and environmental losses due to reactor fire and explosion.

The event tree analysis is used to generate different accident scenarios due to reactor temperature deviations including process near-misses, incidents, and accidents. Then, above mentioned loss categories are assigned to each process end state. For example, process near-misses are assumed to only cause loss of product quality and the process accidents due to loss of material, and energy are assumed to cause all types of loss. The prior monetary evaluation and estimation of these loss categories can be performed using available information from expert knowledge, incident investigation reports, and similar processes. However, when process near misses and incidents do occur, the user perceives new information about the system true loss by analyzing the evident consequences which can be used to revise the basic loss information.

As shown in Table 2.2, the amount of loss associated with overheating of the liquid (temperature above the target) is considered to be more compared to the loss encountered when reactor temperature decreases (temperature below the target). Therefore, a combined function is used for each of the LFs allowing the LFs to have asymmetric shapes. For example, for QLF, the piecewise function is used as $L(T_{reactor},T_T) = B_1 (T_{reactor} - T_T)^2$ for $T_T \leq T_{reactor} < T_{max}$ and $L(T_{reactor},T_T) = B_2 (T_{reactor} - T_T)^2$ for $T_{min} < T_{reactor} < T_{T,}$ respectively.

T_T, T_{min}, and T_{max} are defined in Table 2.2. Similar piecewise functions are used for the other three LFs.

To apply each of the four LFs, the parameters of the LFs (estimated maximum loss and shape parameter) should be determined first. Based on the formulations given in Table 2.1, the QLF and the INLF require only the primary information which is the target reaction temperature and estimated maximum loss values; i.e. T_T, EML_1, and EML_2 in Table 2.2. For the MINLF and IBLF, auxiliary information given in Table 2.2 for the loss values at upper and lower specified reactor temperature limits are used to obtain the shape parameters. The least-squares estimation method is used for this purpose in accordance with Equation 2.3:

$$\min_{SP>0} = \sum_{i=1}^{n} \{L_i - LF_{T_{reactor\_i}}\}^2 \qquad (2.3)$$

where $LF_{T_{reactor\_i}}$ is the value of any of MINLF or IBLF at a given $T_{reactor}$, L_i is the assumed loss values mentioned in Table 2.2, and SP is the shape parameter γ for MINLF and the shape parameter α for IBLF.

All four loss functions are plotted in Figure 2.1, along with the two auxiliary data points which are the assumed amount of system loss at upper specified reactor temperature (set point for high temperature alarm) and lower specified reactor temperature (set point for low temperature alarm); see Table 2.2 for the description and the value of the defined process information. Figure 2.1 shows the flexibility of MINLF and IBLF as it is possible to estimate their shape parameters using the available process information. If more information is available, more accurate shape parameters can be estimated. As shown in

Figure 2.1, QLF and INLF are not flexible as it is not possible to change the shapes of the functions. For this case, QLF underestimates and the INLF overestimates the reactor system loss due to temperature deviations.

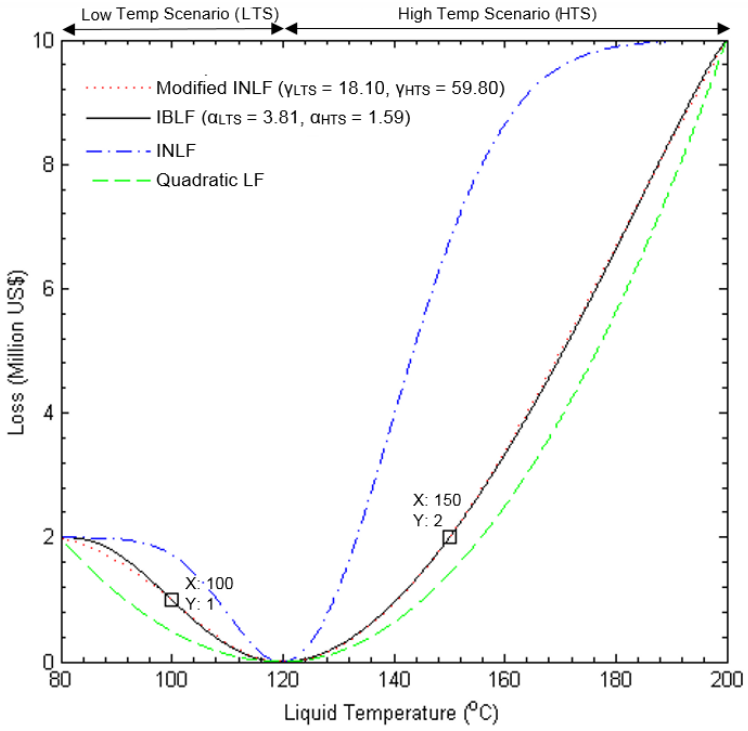


Figure 2.1. Reactor loss using different loss functions

2.6.2. Risk-Based Warning Analysis of the Reactor System

The calculated system loss using LFs is then used in Equation 2.1 to determine the operational risk of the reactor. The proposed operational risk-based method is then used to design the alarm system. To make the case simple, it is assumed that there is only one alert generated from the temperature control system. The probability of occurrence of both high temperature and low temperature scenarios is assumed as 0.01 per year. The process safety

response time for both temperature deviation scenarios is considered as three minutes. These values are obtained from process history and expert knowledge. Figure 2.2 shows the final risk of the alert determined by application of the four LFs in Equation 2.2.

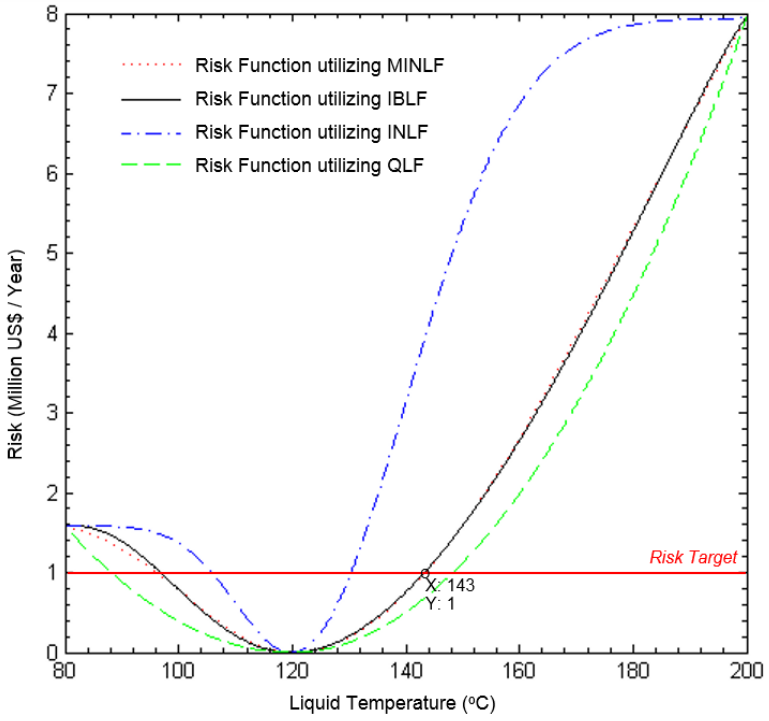


Figure 2.2. Reactor operational risk for different loss functions

When the reactor temperature deviation exceeds the set point for the high alarm (120°C) a warning will be issued. Based on the deviation, the proposed model calculates the real-time risk for the current reactor temperature. Then the estimated risk is compared with the risk target. The operational risk target for the reactor system is considered as 1 million dollars per year; see Figure 2.2. If the estimated risk is less than the target risk, the warning will be classified as alert. If the estimated risk exceeds the target risk, the warning will be

upgraded to alarm and will be annunciated on a different display. The target risk may need to be adjusted for the alarm management system to satisfy the plant safety goals [33].

2.6.3. Sensitivity Analysis

As discussed in Section 2.6.1, the application of the LFs requires determination of the parameters, i.e. the estimated maximum loss (EML) and the shape factor. To study the effect of utilizing different LFs and the effect of uncertainty in the parameters of the LFs on the estimated risk values, a sensitivity analysis is conducted on four LFs. These are: IBLF with $\alpha = 1.59$, INLF with $\gamma = 20$, MINLF with $\gamma = 59.80$, and QLF with $B = 0.0016$. See Table 2.1 and Section 2.2.1 for the list of formulations of each LF and the definition of the parameters, respectively. The selected shape parameters are associated with the LFs shown in Figure 2.1.

The sensitivity analysis results are presented in Figure 2.3. For the case study conducted in this work, it is evident from Figure 2.3(a) that the maximum effect on the estimated risk occurs when the INLF is used followed by QLF. A small deviation in the shape of these two LFs results in a significant change in the estimated risk. The sensitivity of the estimated risk is less for both the MINLF and IBLF. However, the estimated risk is more sensitive for the shape parameter when using IBLF as compared to MINLF. A positive error of about 20% in the shape parameter of MINLF may underestimate the risk by 9.6%, whereas an error of the same magnitude in the shape parameter for the IBLF will overestimate the risk by 45%.

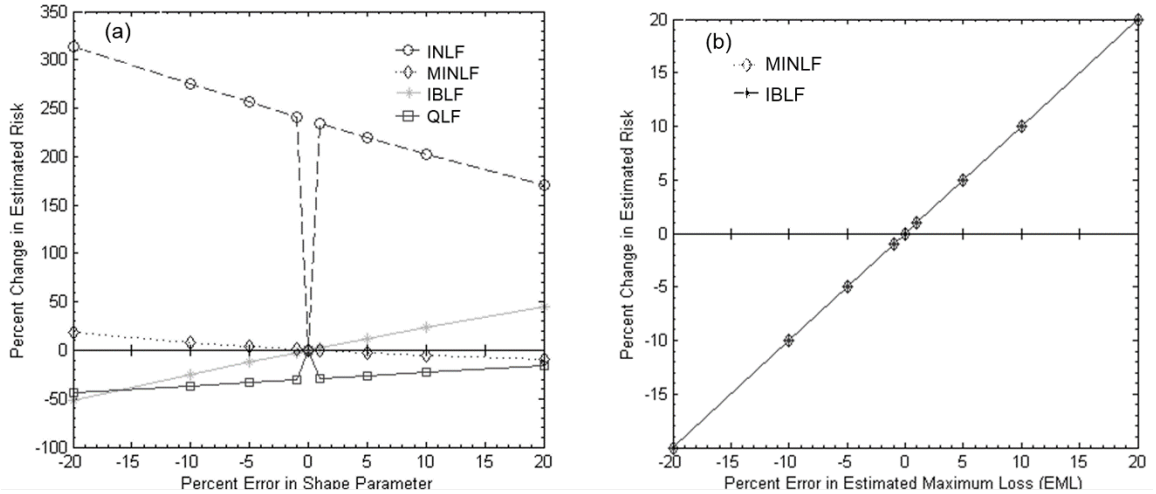


Figure 2.3. Sensitivity analysis: (a) for different EML values, (b) for different shape parameter

MINLF and IBLF have different behavior regarding their impact on the estimated risk as a result of an error in the shape parameter. When using IBLF, positive errors in the shape parameter (higher than the base value) causes positive error in the estimated risk and negative errors (lower than the base value) causes negative error of the same magnitude in the estimated risk. However, as can be seen in Figure 2.3(a), MINLF has opposite behavior as compared to IBLF. Also, when comparing MINLF with IBLF, the estimated risk is more sensitive when there is a negative change in the shape factor as compared to a positive change. A negative error of 20% in the shape factor may overestimate the risk by 18.5%, whereas a positive error of the same magnitude may underestimate the risk by just 9.6%. Figure 2.3(b) shows the sensitivity analysis results related to estimated maximum loss change when using MINLF and IBLF. It is evident from Figure 2.3(b) that the sensitivity of the estimated risk is the same for both IBLF and MINLF. An error of 20% results in an error of the same magnitude in the estimated risk, i.e. a linear dependence.

Overall, it can be concluded from this study that the proper selection of the LF is very important. From the four LFs studied, the MINLF shows comparatively more robust results followed by IBLF. The results of the sensitivity analysis show that a conservative value of the estimated maximum loss should be avoided for both MINLF and IBLF. Considering the results of the sensitivity analysis for the shape factor, it is evident that a conservative value of the shape factor for IBLF should also be avoided. However, it is advisable to use a conservative value of the shape factor for the MINLF.

2.6.4. Comparison of Loss Functions

MINLF and IBLF are found to demonstrate better performance than QLF and INLF. Compared to MINLF, IBLF has an additional advantage of covering asymmetrical cases, i.e. when the target is not at the middle of upper and lower specification limits. However, for the cases when the loss functions should approach different maximum loss values at both sides of the target, like the case study conducted in this work, a piecewise function should be used. Therefore, for asymmetric maximum losses at both sides of the target, which is the applicable case when analyzing most process systems, IBLF and MINLF have the same restriction.

By considering the overall performance of the LFs studied in this work, the MINLF and IBLF could be the optimum choice for depicting the system loss associated with process variations. Choice of LF should be based on the process behavior and availability of loss data. The shape of both MINLF and IBLF can be modified to suit the practitioner's needs for both symmetric and asymmetric problems. However, construction of the MINLF is

easier when compared to IBLF due to simplicity of the formulation and also there is no need for transformation of the scales. Considering the sensitivity analysis results, it is also shown that the MINLF has comparatively more robust performance.

These results are based on the loss data associated with considered reactor in this work. However, it cannot be concluded that a specific loss function is uniformly better than the others in all applications. They are presented as alternative choices for a safety practitioner and they should be used depending on the problem on hand. Choice of LF should be based on the process behavior and availability of loss data.

2.7. Conclusion

Loss functions are used to design an operational risk-based warning system in this work. This is a paradigm shift that will benefit the process industry in terms of continuously improving process safety through proactive loss minimization. Instead of relying on the safety level which has been designed during the system design, the utilization of the proposed approach integrates the safety improvement into the daily activities through loss minimization. By applying the operational risk-based warning system design, the warnings will be differentiated into alerts and alarms. This helps in significant reduction of the generated alarms to meet the recommended alarm regulations. Using the proposed approach, operators can prioritize the alarms and will have more time to make informed decisions. This prevents unnecessary activation of the safety-instrumented systems and minimizes the system loss due to reduced operational and maintenance costs.

The effect of using different loss functions such as QLF, IBLF, INLF, and MINLF are compared and it is concluded that the proper selection of the LF is very important. For the case of the reactor studied here, the MINLF and IBLF are found to demonstrate better performance than QLF and INLF as their profile can be changed to fit the available loss data.

Use of loss functions help to continuously updates the operational risk as per current state of the process. The updated risk could be used as informed decision making variable. The proposed approach helps to foster a zero-loss safety culture and provides the infrastructure for continuous safety improvement. Extending the proposed methodology into multivariate problems is a topic of future research.

2.8. References

[1] Taguchi G. Quality Engineering In Production Systems. First. New York: McGraw-Hill; 1989.

[2] Taguchi G, Clausing D. Robust quality. Harv Bus Rev 1990;68:65–75.

[3] Kapur KC, Cho BR. Economic design of the specification region for multiple quality characteristics. IIE Trans 1996;28:237–48. doi:10.1080/07408179608966270.

[4] Kulkarni SS. Loss-based quality costs and inventory planning: General models and insights. Eur J Oper Res 2008;188:428–49. doi:10.1016/j.ejor.2007.04.041.

[5] Leung BPK, Spiring FA. Some Properties of the Family of Inverted Probability Loss Functions. Qual Technol Quant Manag 2004;1:125–47.

[6] Spiring FA. The reflected normal loss function. Can J Stat 1993;21:321–30.

[7] Pan J, Chen S. A new approach for assessing the correlated risk. Ind Manag Data Syst 2012;112:1348–65. doi:10.1108/02635571211278965.

[8] Sii HS, Ruxton T, Wang J. Taguchi concepts and their applications in marine and offshore. J Eng Des 2001;12:331–58. doi:10.1080/0954482011008594 0.

[9] Ordoobadi S. Application of Taguchi loss functions for supplier selection. Supply Chain Manag An Int J 2009;14:22–30. doi:10.1108/13598540910927278.

[10] Dey S. Bayesian Estimation of the Parameter and Reliability Function of an Inverse Rayleigh Distribution. Malaysian J Math Sci 2012;6:113–24.

[11] Leung BPK, Spiring FA. The inverted beta loss function: properties and applications. IIE Trans

2002;34:1101–9.

[12] Murphy TE, Tsui KL, Allen JK. A review of robust design methods for multiple responses. Res Eng Des 2004;15:201–15. doi:10.1007/s00163-004-0054-8.

[13] Fathi Y, Poonthanomsook C. A Quartic Quality Loss Function and Its Properties. J Ind Syst Eng 2007;1:8–22.

[14] Chan WM, Ibrahim RN, Lochert PB. Quality evaluation model using loss function for multiple S-type quality characteristics. Int J Adv Manuf Technol 2004;26:98–101. doi:10.1007/s00170-003-1980-8.

[15] Pan J, Pan J. A Comparative Study of Various Loss Functions in the Economic Tolerance Design. 2006 IEEE Int. Conf. Manag. Innov. Technol., vol. 2, Singapore: Ieee; 2006, p. 783–7.

[16] Joseph VR. Quality Loss Functions for Nonnegative Variables and Their Applications. J Qual Technol 2004;36:129–38.

[17] Sun F-B, Laramee J-Y, Ramber JS. On Spiring's normal loss function. Can J Stat 1996;24:241–9.

[18] Spiring FA, Yeung AS. A general class of loss functions with industrial applications. J Qual Technol 1998;30:152–60.

[19] Chan WM, Ibrahim RN. Evaluating the quality level of a product with multiple quality characteristics. Int J Adv Manuf Technol 2004;24:738–42. doi:10.1007/s00170-003-1751-6.

[20] Hsu CM. Solving Multi-Response Problems Through Neural Networks and Principal Component. J Chinese Inst Ind Eng 2001;18:47–54. doi:10.1080/10170660109509504.

[21] Suhr R, Batson RG. Constrained Multivariate Loss Function Minimization. Qual Eng 2001;13:475–83. doi:10.1080/08982110108918676.

[22] Pignatiello JJ. Strategies for robust multiresponse quality engineering. IIE Trans 1993;25:5–15. doi:10.1080/07408179308964286.

[23] Artiles-León N. A pragmatic approach to multi-response problems using loss functions. Qual Eng 1996;9:213–20. doi:10.1080/08982119608919037.

[24] Ma Y, Zhao F. An Improved Multivariate Loss Function Approach to Optimization. J Syst Sci Syst Eng 2004;13:318–25.

[25] Su C, Tong L. Multi-response robust design by principal component analysis. Total Qual Manag 1997;8:409–16. doi:10.1080/0954412979415.

[26] Chan WM, Ibrahim RN, Lochert PB. Evaluating the product quality level under multiple L-type quality characteristics. Int J Adv Manuf Technol 2004;27:90–5. doi:10.1007/s00170-004-2158-8.

[27] Abdul-Kader W, Ganjavi O, Solaiman A. An integrated model for optimisation of production and quality costs. Int J Prod Res 2010;48:7357–70. doi:10.1080/00207540903382881.

[28] Raiman L, Case K. Monitoring continuous improvement using the Taguchi loss function. Int. Conf. Ind. Eng., San Francisco, California: 1990, p. 391–6.

[29] Tong LI, Su CT, Wang C-H. The optimization of multiresponse problems in the Taguchi method. Int J Qual Reliab Manag 1997;14:367–80.

[30] Tong LI, Su CT. Optimizing multi-response problems in the Taguchi method by Fuzzy multiple attribute decision making. Qual Reliab Eng Int 1997;13:25–34.

[31] Pan JN. A new loss function-based method for evaluating manufacturing and environmental risks. Int J Qual Reliab Manag 2007;24:861–87. doi:10.1108/02656710710817135.

[32] AIChE. Guidelines for Risk Based Process Safety. First. New York: Wiley; 2007.

[33] Chang Y, Khan F, Ahmed S. A risk-based approach to design warning system for processing facilities. Process Saf Environ Prot 2011;89:310–6. doi:10.1016/j.psep.2011.06.003.

[34] Ahmed S, Gabbar HA, Chang Y, Khan FI. Risk based alarm design: A systems approach. 2011 Int. Symp. Adv. Control Ind. Process. ADCONIP 2011, Hangzhou, China: 2011, p. 42–7.

[35] Woods DD. The alarm problem and directed attention in dynamic fault management. Ergonomics 1995;23:2371–93.

[36] EEMUA. Alarm Systems: A Guide to Design, Management and Procurement. 2nd ed. EEMUA; 2007.

3. RISK-BASED OPERATIONAL PERFORMANCE ANALYSIS USING LOSS FUNCTIONS[2]

Preface

A version of this manuscript has been published in the Journal of Chemical Engineering Science. I am the primary author of this paper. Along with the co-authors, Faisal Khan and Salim Ahmed, I developed the conceptual model and subsequently translated this to a numerical risk assessment model using loss functions. I carried out most of the literature review, data collection and analysis. I prepared the first draft of the manuscript and subsequently revised the manuscript based on the co-authors' feedback and also the peer review process. The co-author Faisal Khan helped in developing the concepts/models and their testing, reviewed and corrected the models and results, and contributed in preparing, reviewing and revising the manuscript. The co-author Salim Ahmed contributed through support in the development, testing and improvement of the model. Salim Ahmed also assisted in reviewing and revising the manuscript.

Abstract

This paper proposes a risk-based process performance assessment methodology using loss functions. The proposed method helps to overcome the existing challenges in assessing impacts of deviations of process variables on safety and economy of a process operation.

[2] Hashemi et al. Chemical Engineering Science 2014;116:99–108.

The inverted Beta loss function is used to incorporate the effects of process deviations on the safety and quality losses. The demand rate adjustment factor is used to model the effect of process deviations on the failure probability of safety systems. The probability of a failed process state due to abnormal events is continuously updated based on the current value of the characteristic variables. The use of the loss function approach in combination with probability updating provides a continuously revised risk estimation. Such a real-time risk profile provides a leading performance indicator for decision-making at an operational level. As an example, a temperature surge in a continuous stirred tank reactor is used to demonstrate the efficacy of the proposed methodology.

Keywords: Loss function; Safety loss; Quality loss; Performance analysis; Risk.

3.1. Introduction

Meeting the financial targets of stakeholders, as well as the quality requirements of consumers without compromising operational safety is critical for the economical and safe operation of a process facility. Achieving this goal requires the proper management of the process facility's operational performance along with management of process safety. An efficient management of an operation demands a leading performance indicator [1] that focuses on both safety and quality elements. As mentioned by the United States Center for Chemical Process Safety (CCPS), "facilities should monitor the real-time performance of management system activities rather than wait for accidents to happen. Such performance monitoring allows problems to be identified and corrective actions to be taken before a serious incident occurs" [2].

A shortcoming of traditional process safety management (PSM) systems is that the safety program is not well integrated with other functions of an organization [3]. Here, an organization is defined as a multidisciplinary entity having a collective goal of economical and safe operations of a process facility. This paper focuses on the integration of safety analysis with operational performance analysis, and proposes a method to assess quality and safety losses simultaneously. Through such integration, quality management strategies can be adopted to improve process safety, and vice versa. In order to achieve the highest levels of safety and quality, with the ultimate goal to foster a zero-incident and zero-defect culture, the aim should be to eliminate the main sources of the losses, i.e. process deviations. For process facilities the causes of deviations may include feed variability, mechanical and operational integrity degradation, wrong setting, and improper methods.

To analyze the impact of process deviations on safety, Hashemi et al. [4] proposed the application of loss functions and compared their properties. The first contribution of this paper is to expand the earlier work of Hashemi et al. [4] by integrating both safety and quality losses associated with process deviations. Secondly, the concept of operational risk is introduced. The operational risk is estimated by combining the loss function-based consequence assessment method with a probabilistic approach to define the probability of undesired process states. These two contributions enable a novel risk-based process performance assessment methodology that continuously revises the risk based on the current process state. The developed risk profile is a leading indicator of process performance that can be used for day-to-day operational decision-making.

The motivations behind the development of an integrated framework to combine safety and quality losses, and the use of risk as a performance indicator, are illustrated in Section 3.2. The developed methodology is outlined in Section 3.3 followed by a case study in Section 3.4. Finally, the concluding remarks are presented in Section 3.5.

3.2. An Overview and Motivations

3.2.1. Integration of Safety and Quality Management

Continuous improvement is the key to quality management. Taguchi [5] suggested the application of a quadratic loss function for quality loss modeling to promote the concept of continuous improvement. According to Taguchi's philosophy [5], a process imparts zero quality loss only if the quality characteristic of interest is at its target. Even a small deviation from the target imposes loss to society [5]. Here, society is the combination of the producer, the consumer, and the environment. As reviewed by [4], after Taguchi [5], several researchers proposed different types of loss functions to overcome the limitations and inflexibility of the quadratic loss function.

Deviations of characteristic variables also affect operational safety. Process deviations, along with failures of the control systems and the layers of protection, result in safety losses. The more the process safety characteristic deviates from its target operating conditions, the larger the demand for the activation of corresponding control and safety systems. Due to inevitable uncertainty in the activation and successful operation of these systems, any deviation of the process characteristic increases the possibility of incurring

loss. Thus, there is inherent similarity between safety loss and Taguchi's quality loss model. Therefore, loss functions can also be used to model safety losses of a process.

Hashemi et al. [4] compared the properties of different loss functions in modeling system loss and reviewed their applications in process safety analysis. The modified inverted normal loss function (MINLF) and the inverted Beta loss function (IBLF) were found to be more adaptable to depict system safety losses associated with process variations [4]. This work aims to further expand the application of loss functions for integrated modeling of both safety and quality losses due to the variability of process operations.

The benefits of integrating safety and quality management systems have been discussed in the literature [3,6]. While quality improvement methods strive to minimize the variability inherent in product quality, safety management strives to minimize the chance of occurrence and the severity of incidents that can cause loss [7,8]. The IBLF is used in this work for integrated modeling of safety and quality losses due to process deviations.

3.2.2. Risk-Based Process Performance Assessment

Combining loss models with the probability of process deviations provides a framework to develop a risk-based approach to process safety and quality performance assessment. As risk includes both the probability of an end process state and its consequences, a risk-based approach reduces the potential for assigning an undue amount of resources to manage lower-risk events, thereby freeing up resources for tasks that address higher-risks [2,9].

To be useful for operational decision-making, risk estimation has to be updated with variations in a process. As reviewed by [10] and [11], there have been efforts to make risk

assessment methods dynamically adaptable with real-time changes occurring in a process. Generally, risk is updated based on the number of events recorded over time [12]. Existing literature on this topic focuses mainly on probability updating. However, the effect of process deviations on the real-time value of system loss is not taken into account. By utilizing a loss function approach to model both safety and quality losses associated with undesired process end states, this work proposes an integrated risk-based performance analysis methodology. The effects of process deviations on both system loss and on the probability of failure of safety barriers are considered.

3.3. Methodology

3.3.1. Identification of the Key Process Characteristics

The first step in the development of a process performance assessment methodology is the identification of the key process characteristics. A key characteristic is a feature that, if non conforming, missing, or degraded, may cause unsafe conditions and/or a loss of product quality. For example, operating temperature is the key process characteristic of a polymerization reactor. Different approaches, such as check lists, preliminary hazard analysis (PHA), failure modes and effects analysis (FMEA), fault tree analysis (FTA), hazard and operability study (HAZOP), and master logic diagrams [13], are often used to identify key characteristics. This work assumes that a univariate key process characteristic can be assigned to a system. The consideration of systems with multiple characteristics is considered in Chapter 7.

3.3.2. Scenario Analysis

Once the key process characteristic is identification of potential scenarios associated with possible deviations. The outcome scenario, for example an accident, may result from a single event or combinations of events [13]. Based on the type of process upset and the performance of control and safety barriers, the scenario could involve quality loss, safety loss, or both. The maximum credible accident scenario analysis method developed by Khan and Abbasi [13] can be used as a criterion to identify credible scenarios among a large number of possibilities.

When analyzing accident scenarios, the first step is the identification of abnormal situations, which are process deviations in the current context. Deviations result from external factors or the ineffectiveness of the control systems in rejecting disturbances. The deviation of a key process characteristic, along with the failure of protection systems, will cause undesirable process end states. Table 3.1 represents the generic categorizations of process end states and gives examples from both safety [14] and quality points of view. The next step involves identification of the sequence of events which transforms an abnormal situation into its process end state. Event tree analysis is used to analyze the influence of the failure or success of different control and/or safety systems when deviations occur. Having envisaged the accident scenarios, as discussed in the next section, event consequences should be identified and quantified.

Table 3.1. Categorization and definition of event outcomes, adapted from [14]

Outcome	Definition	Process safety examples	Process quality examples
Near-miss	An event stemming from process deviation that does not result in an actual safety loss but has the potential to do so	Unnecessary activation of the trip system; release with no impact on people and property	Minor quality loss due to deviation of the key process characteristic from the target
Mishap	An event or sequence of events that causes minor to moderate impact on people, property, the environment the environment, or the final product	Minor health effects; minor impact on property and the environment	Minor health effects; minor impact on property and specifications and material wastage
Incident	An event that causes considerable harm or loss to people, property, the environment, or the final product	Temporary disability or permanent minor disability; localized damage to assets and the environment	Considerable internal losses due to rework and scrapped material as well as external losses such as warranties and returned products
Accident	An event that causes serious impact to people and assets. An event like this has heavy financial loss and receives national media attention	One or more fatalities or permanent major disabilities. Process shutdown and heavy financial losses	Heavy internal and external losses
Disaster	An event that causes multiple serious losses. Such an event receives international media attention	Multiple fatalities and extensive damage to assets and production. It may cause a shutdown of the plant for a significant time period, possibly forever	Heavy internal and external losses. It may cause a loss of market share for a significant time period. This condition is referred to as a quality meltdown (QM)

3.3.3. Consequence Assessment

3.3.3.1. Loss Modeling

In traditional process loss modeling methods, no loss with respect to product quality and system safety is assumed until the key process characteristic deviates from the operational boundary. Using the loss functions, consequences can be assessed by any deviation of the process characteristics from its target; zero loss is only considered for on-target operation. This work uses the IBLF to model system loss, as it is easy to construct and its shape can

be modified to suit the practitioner's needs for both symmetric and asymmetric losses [4]. Modeling the system loss using loss functions requires identification of the maximum loss and determination of the shape parameters. In this study, the following two steps are proposed to model the system loss:

- Estimation of maximum loss: According to Marsh Risk Consulting, the estimated maximum loss (EML) is defined as "the loss that could be sustained under abnormal conditions with the failure of all protective systems" [15].
- Identification of shape parameter: A non-linear search approach based on the least-squares method is proposed to determine the shape parameter for attaining a suitable fit to the actual loss.

3.3.3.2. Identification and Estimation of Losses

There is a general agreement in the literature about four major categories of the potential losses for any given scenario. These losses are production loss (PL), asset loss (AL), human health loss (HHL), and environmental cleanup cost (ECC) [16,17]. In addition to these loss categories, the deviation of key process characteristics could result in considerable quality loss. There have been some efforts toward the integration of quality and safety aspects due to process upsets, mainly in the manufacturing industry [3,18]. Pariyani et al. [19] considered quality loss in their proposed risk analysis methodology for process industries. In this work, quality loss is considered as one of the major loss categories resulting from process variations. Loss categories with their cost indicators are shown in Figure 3.1.

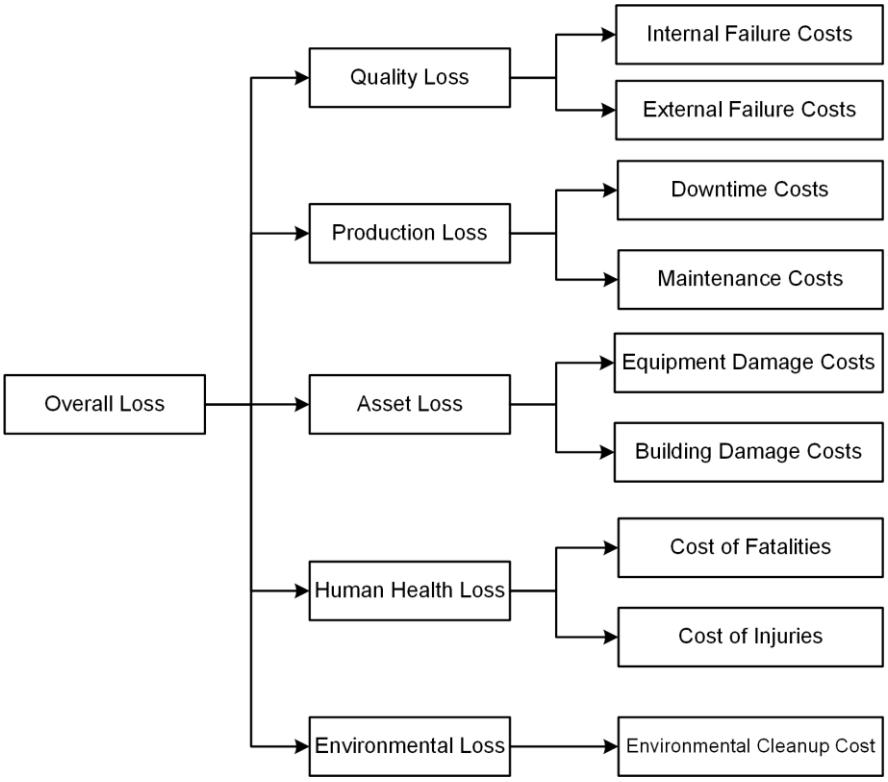


Figure 3.1. Loss categories with their loss indicators (adapted from [16] and [20])

To estimate maximum PL, AL, HHL, and ECC, the process downtime and affected areas should be calculated using dispersion models discussed in CCPS [21], Khan and Amyotte [16], and American Petroleum Institute [17]. The estimation of maximum quality loss requires a different approach. Traditionally, quality costs are classified into four categories [22]:

- Appraisal costs: Typical examples are inspection, quality audits and acceptance tests
- Prevention costs: Maintenance of equipment, training operators and improving procedures, etc.

- Internal costs: Scrap, replacement, rework repair, etc.
- External costs: Examples are warranties, loss of market share, and loss of reputation

Internal and external failure costs arise when the failure is detected inside of the organization or by the customer [23]. The possibility of experiencing internal and external failure costs prompts companies to incur appraisal and prevention costs. Therefore, the maximum quality loss can be considered as the summation of the maximum internal and external failure costs. Margavio et al. [23] discussed quality costs using a modern accounting approach, which can be used as a guideline when estimating the maximum quality loss. Quantification of these loss categories is based on the worst case conditions to obtain the estimated maximum loss for each category in Figure 3.1. Finally, the total estimated maximum loss, *EML*, for each abnormal event can be obtained from

$$EML_j = \sum_{i=1}^{5} EML_{i,j} \tag{3.1}$$

where *j* denotes the number of undesired event scenarios and *i* counts the number of different losses. Having determined the estimated maximum losses, the next step is the development of the loss function. 3.3.3.

3.3.3.3. Loss Function Development

Among the five main loss categories shown in Figure 3.1, the asset loss, environmental cleanup cost, and human health loss usually occur instantaneously in the case of an accident. Therefore, the associated loss function has an almost step-like shape and the system incurs these losses after the key process characteristic deviates beyond a safe

operating boundary. However, for the case of production loss and quality loss, the system might experience these losses as soon as the key characteristic deviates from the target operating value. This work uses IBLF to model these loss categories, as its shape is flexible enough to change from a step-like loss function to a smoothly increasing loss function. However, from a practical point of view, different types of loss functions may be required to model each loss category. The review of the properties of different loss functions and the sensitivity analysis approach in Hashemi et al. [4] can be used to select a suitable loss function for specific applications.

As mentioned above, this study uses the IBLF developed by Leung and Spiring [24] to quantify losses. Let $f(x) = 1/B(\alpha_b, \beta_b)x^{\alpha_b - 1}(1-x)^{\beta_b - 1}$ be the standard Beta probability density function (PDF), with unique maximum at $x = (\alpha_b - 1)/(\alpha_b + \beta_b - 2)$ for $\alpha_b > 0$ and $\beta_b > 0$, and let the target be $T = (\alpha_b - 1)/(\alpha_b + \beta_b - 2)$. Using the unique maximum conditions associated with the Beta distribution, a linear relationship can be established between α_b and β_b through T. This relationship can be written as $\alpha_b - 1 = T(\beta_b - 1)/(1-T)$. The loss function formulation associated with inverting the Beta PDF is:

$$IBLF(x,T) = EML\{1 - C\left[x(1-x)^{(1-T)/T}\right]^{(\alpha_b - 1)}\} \qquad (3.2)$$

where $C = [T(1-T)^{1-T/T}]^{1-\alpha_b}$, x is the process characteristic, and T is the process target. *EML* is the estimated maximum loss determined from Section 3.3.3.2. In Equation (3.2), α_b is the shape parameter which adjusts the penalty for deviation from the target. A large α_b indicates that the process can tolerate relatively small deviation. The derivation of the

above relationships and the discussion of the properties of IBLF compared to other types of loss functions can be found in [24] and [4] respectively.

Determining the shape parameter of the IBLF is the next step. Loss functions are chosen to reflect the loss associated with process deviations. In many cases only partial information regarding the actual loss associated with a deviation from target, *T*, is known. The most frequent case is when the maximum loss (and its first occurrence) is known and the loss at target is assumed to be zero. This is referred to as the primary loss information. If additional information is available, it can be used to provide a better representation of the loss function, keeping in mind that the goal is to accurately depict the losses associated with deviations from the target [25].

The primary loss information is denoted by $[KPC_T, 0]$ and $[KPC_T + \Delta_j, EML_j]$ where KPC_T is the key process characteristic at its target value, EML_j is the estimated maximum loss as discussed in Section 3.3.3.2, Δ_j is the distance from the target to the point where EML_j occurs, and j denotes the number of undesired event scenarios. For example, for the case where internal pressure is the key process characteristic for an undamaged piece of equipment, API 581 standard assumes that the probability of loss of containment will equal 1.0 when the overpressure is equal to 4 times the maximum allowable working pressure [17]. Similarly, when enough information is not available, Δ can be considered as being 4 times larger than the maximum operating value of the process characteristic.

Having defined the primary loss information, and losses at a set of n additional points, i.e. $\{[x_1, L_1], [x_2, L_2], \ldots, [x_n, L_n]\}$, one can determine the shape parameter α_b by applying a non-linear search procedure, such as the least squares method

$$\min_{\alpha>0} = \sum_{i=1}^{n} \{L_i - IBLF_{x_i}\}^2 \quad (3.3)$$

where i counts the number of data points, L_i is the loss at each data point, and $IBLF_{xi}$ is the value of IBLF at x_i. Lacking secondary information, a pragmatic choice is to set α_b to 1.03. This value is determined from back calculation of Equation (3.2). In this case the loss is about 50% of the estimated maximum loss when a process characteristic is at distance Δ from the target. The resulting loss function corresponds to the step loss function in the same situation.

3.3.4. Analysis of Scenario Probability

A major concern with safety barriers is their failure during an emergency situation. This can ultimately cause failure of the concerned equipment. Conventional probability analysis methods are widely used in risk-based process safety management. These methods use generic failure data and are static in nature [12]. Dynamic probability analysis methods have been used to update risk as new evidence becomes available. However, most of the existing dynamic risk assessment approaches consider the effects of accident precursor data on the probability of safety barriers failure.

This work proposes an approach, which considers the effects of process deviations on the failure probabilities of safety barriers. The basic principle is to adjust the abnormal event (process deviation) frequency using the demand rate adjustment factor (DRAF). The adjusted demand rate is then multiplied by the probability of failure on demand (PFD) of safety barriers to calculate the failure rate. Finally, the calculated failure rate is used in

event tree analysis to determine the probability of process end states. The approach proposed in this work and the probability adapting method proposed by Khakzad et al. [12] can be integrated to consider the parallel effects of accident precursor data and process deviations on the failure probability of safety barriers.

3.3.4.1. Estimation of the Failure Probability of a Safety Barrier

The failure rate of safety systems refers to their failure to operate during emergencies. The demand rate placed on a safety barrier and the probability of failure on demand are considered in the estimation of failure rate, FR:

$$FR_{SB_i}(x,T) = DR(x,T) \times PFD_{SB_i} \tag{3.4}$$

where x is the process characteristic, T is the process target, DR is the demand rate (demands or events/year) placed on the device, and PFD_{SB_i} is the probability of failure on demand of safety barrier i (failure/demand).

There are some industry-specific data that provide generic values of PFD. For example, API 581 includes PFD values for pressure relief devices expressed as Weibull curves [17]. Regarding the safety instrumented systems (SIS), ISA-TR84.00.02 [26] includes simplified equations to calculate total PFD by calculating and adding the PFD of individual components in each safety instrumented function (SIF). The user may obtain the PFD of SIS components from the vendor for the actual functional test interval. Considering the operator intervention as a safety barrier, the PFD of the operator can be determined using human error probability (HEP) assessment technique developed recently by Musharraf et al. [27] which uses a Bayesian approach integrated with evidence theory. The present work

assumed that the PFD is invariant with time to simplify the analysis, and a constant PFD for each safety barrier is used in failure rate calculations.

3.3.4.2. Demand Rate Estimation

The demand rate placed on safety devices is dependent on the abnormal event frequency. However, the actual demand rate on a specific safety device is not necessarily equal to the abnormal event frequency. API 581 uses a constant demand rate reduction factor to take credit for additional safety barriers in reducing the actual demand rate on pressure relief devices [17]. This work recognizes that the higher process deviation results in higher demand for the activation of safety barriers. The concept of a demand rate adjustment factor (DRAF) is introduced to account for the difference between the process deviation frequency and the demand rate placed on safety barriers. Accordingly, the demand rate is calculated as the product of process deviation frequency (DF), and DRAF:

$$DR(x,T) = DF \times DRAF(x,T). \tag{3.5}$$

Process DF is a non-negative integer-value and is modelled using the Poisson distribution [14]. DRAF is considered as a function of the process characteristic and is represented by the inverted Beta probability density function in accordance with

$$DRAF(x,T) = 1 - C[x(1-x)^{(1-T)/T}]^{(\alpha_{DRAF}-1)} \tag{3.6}$$

where $C = [T(1-T)^{1-T/T}]^{1-\alpha_{DRAF}}$ and T is the process target. In Equation (3.6), α_{DRAF} is the shape parameter and can be obtained using the procedure described in Section 3.3.3.3, given available information on the performance of safety barriers. The required information may be obtained through a layer of protection analysis (LOPA) [28].

According to Equation (3.6), the DRAF varies between 0 and 1 to recognize the fact that the demand rate on the safety devices is often less than the abnormal event frequency. As an example, trip systems rarely activate during an overpressure demand case. This is due to the operation of other layers of protection, such as control systems and operator intervention, that reduce the likelihood of reaching the trip system's set point. However, upon increasing deviation of the process characteristic and failure of the layers of protection, the actual demand rate placed on safety barriers reaches the process deviation frequency.

3.3.4.3. Estimation of the Frequency of End States

Knowing the failure rate of different safety barriers, the frequency of each process end state can be calculated. Based on the event tree analysis, the frequency of severity level *k*, denoted by $F(C_k)$, can be obtained using

$$F(C_k) = DR \times \sum_m \left[\prod_{SB_i} PFD_{SB_i}^{\theta_{i,k}} \left(1 - PFD_{SB_i}\right)^{1-\theta_{i,k}} \right] \tag{3.7}$$

where *DR* is the demand rate as a function of the process characteristic calculated using Equation (3.5), m represents the number of pathways into sub-events of each end event, and SB_k denotes the safety barrier associated with the level *k*. $\theta_{i,k} = 1$ if the level *k* failure passes the down-branch of the event tree (failure) associated with safety barrier *i*, and $\theta_{i,k} = 0$ if the level *k* failure passes the up-branch (success) of safety barrier *i*. The summation sign in Equation (3.7) is used because $P(C_k)$ is the summation of frequencies of all different scenarios (pathways in the event tree) that result in the occurrence of end state C_k.

3.3.5. Risk Estimation

The calculation of risk for a severity level *k* as a function of process deviation is obtained by combining the frequency and the consequence of each process end state:

$$Risk_k = Loss_k(x,T) \times F_{C_k}(x,T). \tag{3.8}$$

Note that the risk in Equation (3.8) is the operational risk of process deviation outcomes, and has a value of "zero" when the process characteristic is at its target. To determine the overall risk of a system, risk sources other than operational risk should also be evaluated. The overall methodology for operational risk-based process safety and quality performance analysis is summarized in Figure 3.2. The application of the developed procedure is demonstrated in the next section.

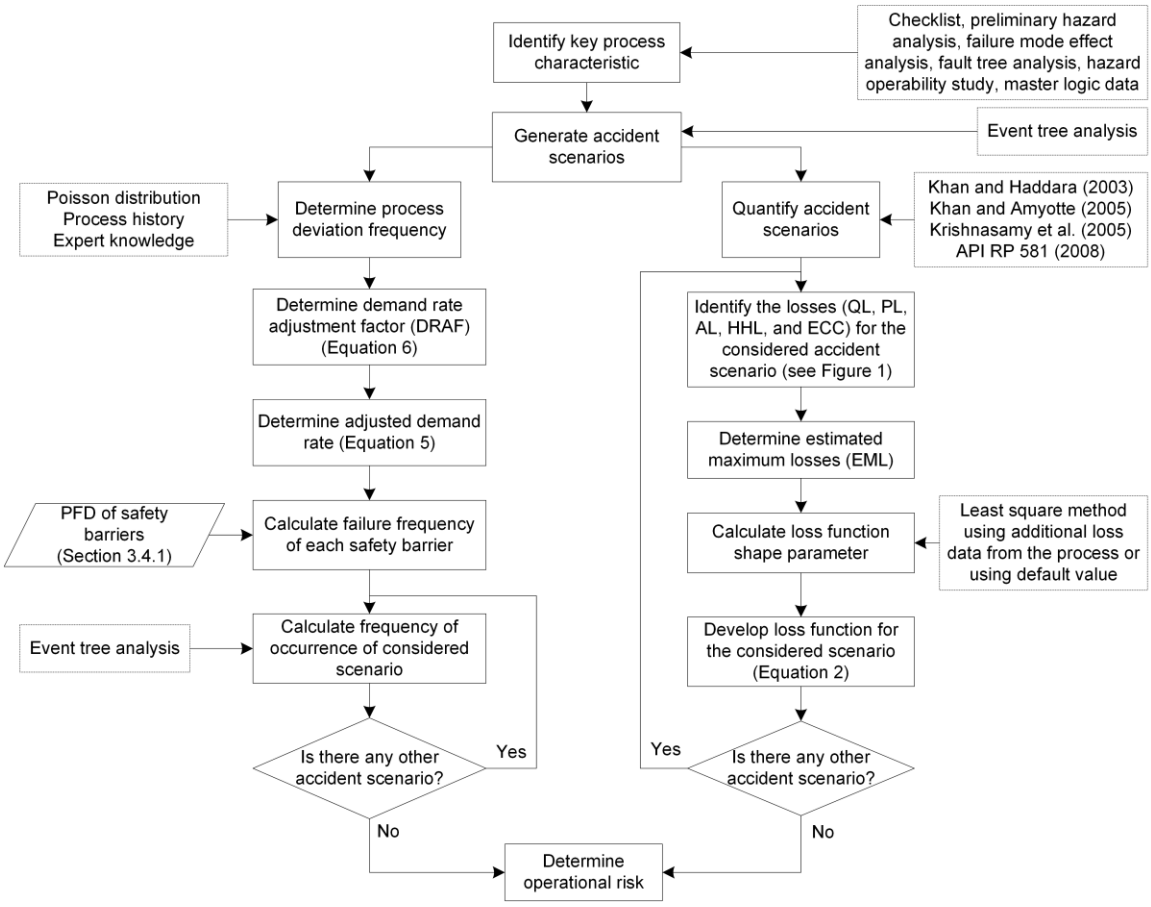


Figure 3.2. Operational risk-based process safety and quality performance analysis methodology

3.4. Case Study: Continuous Stirred Tank Reactor

A temperature deviation event in a continuous stirred tank reactor (CSTR) is taken as a case study to illustrate the proposed procedure. Figure 3.3 shows a schematic of the CSTR with associated protection layers. A description of the control and safety systems of the CSTR can be found in Willis [29] and Meel and Seider [30], respectively. The reactor is fed by a stream rich in reactant *A* of concentration $CA_{(in)}$ and flow rate $F_{(in)}$. Within the system, the following irreversible exothermic reaction takes place $A \rightarrow B \rightarrow C$. The target

reaction temperature is 150 ºC, at which reactant *A* is converted to product *B*. However, at high temperatures *B* undergoes a further reaction and is transformed into undesired by-product *C*, imposing loss on the system. On the other hand, at low temperatures, the reaction cannot take place and product *B* cannot be produced, which results in a loss of product. Moreover, high temperatures may lead to a runaway condition for the reactor. Therefore, the reactor temperature, T_{react}, is identified as the key process characteristic, as it has a direct effect on process safety as well as on the reaction rate.

A coolant stream and a heat exchanger are used to cool the reactor, as shown in Figure 3.3. The objective is to maintain the temperature inside the reactor at the desired value when subjected to changes in inlet concentration (C_{in}) and temperature (T_{in}). When the CSTR is in operation, the conversion and the temperature may undergo large variations in response to disturbances in one or more input parameters. Temperature deviation above the target reaction temperature results in highly exothermic conditions and this situation might be characterised as runaway. The reactor might be uncontrolled at this condition due to autocatalytic decomposition of the reaction and this could result in an explosion. The released material would be toxic, corrosive, and flammable. If the temperature drops below the target reaction temperature, the system experiences quality loss due to off-specification of the final product. This situation can eventually result in reactor shutdown if the temperature drops below the onset temperature of the reaction.

The event-tree model associated with the high-temperature event in the CSTR is shown in Figure 3.4, where P_i, $i = 1, 2, ..., 6$, represent the probability of individual barrier failure. The disturbances in inlet composition and temperature are measured and passed to a feed

forward controller that calculates the necessary coolant flow rate to compensate for any temperature change. If the control system is successful, the system returns to normal operation, denoted by C1-SAFE. When the control system fails to keep the reactor temperature within the safe operating window, the abnormal event propagates through different branches of the event tree based on success or failure of the safety systems. As shown in Figure 3.4, depending on different pathways in the event tree, any of the following outcomes could happen: near- miss (C2-NM), process shutdown (C3-SD), or runaway reaction/quality meltdown (C4-RA/QM). Table 3.2 shows different process end states associated with the high-temperature event in the CSTR.

The control and safety systems for the low temperature event have the same logic as the high temperature event, and can ultimately result in process shutdown and loss of production.

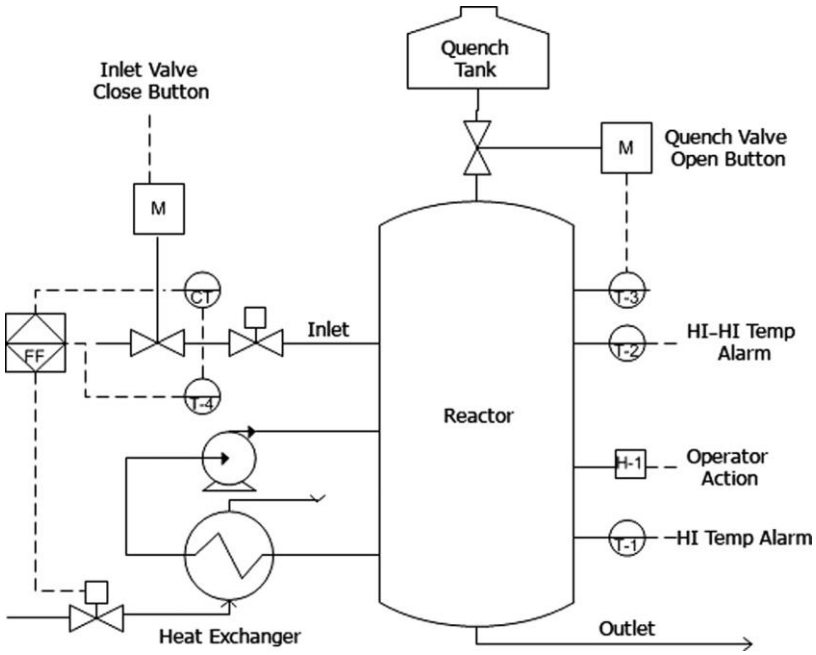


Figure 3.3. Exothermic CSTR and associated protection layers

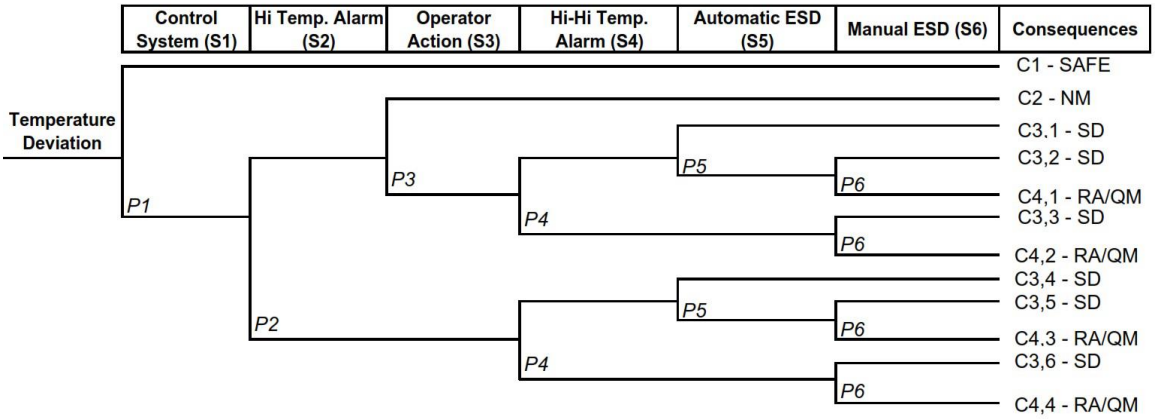


Figure 3.4. Event tree for high temperature event in CSTR

Table 3.2. Process end states associated with the high-temperature event in the CSTR

Process end state	Symbol
Normal operation	C1-SAFE
Near-miss (continued operation)	C2-NM
Process shutdown	C3-SD
Runaway reaction/quality meltdown	C4-RA/QM

3.4.1. Identification of Losses for the CSTR

Table 3.3 represents the estimated maximum values for different loss categories in the high temperature event and the low temperature event in the CSTR. The numbers in Table 3.3 are for illustrative purposes only. In the case of the high temperature event, as shown in Figure 3.4, there are four different process end states due to the high temperature event scenario.

Table 3.3. Estimated maximum losses for different losses for HTE and LTE in the CSTR

Identified EML	EML_{QL}		EML_{PL}		EML_{AL}		EML_{HHL}		EML_{ECC}		EML_{Total}	
Scenario	**HTE**	**LTE**	**HTE**	**LTE**	**HTE**	**LTE**	**HTE**	**LTE**	**HTE**	**LTE**	**HTE**	**LTE**
Loss (10^6 USD)	0.4	0.4	3.6	3.6	20	0	20	0	6	0	50	4

Note: HTE: high temperature event; LTE: low temperature event.

Zero loss is assumed for normal operation (C1-SAFE). For the case of a process near-miss (C2-NM), as shown in Table 3.3, the maximum associated quality loss is estimated as 0.4 million US dollars. The maximum loss associated with process shutdown (C3-SD) is considered as the summation of maximum quality loss (0.4 million dollars) and maximum production loss (3.6 million dollars). The worst case scenario for the high temperature event is the runaway reaction, with the summation of human health loss (20 million dollars), environmental cleanup cost (6 million dollars), asset loss (20 million dollars), production loss (3.6 million dollars) as well as quality loss (0.4 million dollars). It is assumed that the low temperature event will result in a process shutdown and production loss with associated quality loss. No asset loss, human health loss, or environmental cleanup costs are considered for the low temperature event. The estimated maximum losses for all loss categories are obtained in accordance with Section 3.3.3.2. The total estimated maximum loss for each temperature deviation scenario (i.e. low temperature event and high temperature event) is shown in Table 3.3.

3.4.2. Loss Modelling

The IBLF is used to model the CSTR loss due to temperature deviations. The amount of loss associated with a runaway scenario is considered to be greater than the loss encountered when reactor temperature decreases. Therefore, Equation (3.2) is modified to have an asymmetric shape:

$$IBLF(T_{react},T_T)=\begin{cases} EML_{LTE}\{1-C_{LTE}\left[T_{react}(1-T_{react})^{(1-T_T)/T_T}\right]^{(\alpha_{LTE}-1)}\} & T_{react}<T_T \\ EML_{HTE}\{1-C_{HTE}\left[T_{react}(1-T_{react})^{(1-T_T)/T_T}\right]^{(\alpha_{HTE}-1)}\} & T_{react}\geq T_T \end{cases} \quad (3.9)$$

where $IBLF(T_{react},T_T)$ is the overall loss function, $C_j=[T_T(1-T_T)^{1-T_T/T_T}]^{1-\alpha_j}$, EML_j is calculated according to Section 3.3.3.3 with the values shown in Table 3.3, α_j is the shape parameter to be determined using Equation (3.3), T_T is the reactor target (normal) operating temperature, and j is either the high temperature event or the low temperature event scenario.

The next step involves determining the shape parameters for each of the two functions in Equation (3.9). For the high temperature event scenario, the primary loss information includes $[150,0]$ and $[220,500]$ where the first numbers in each dataset are the reactor temperature in degrees centigrade and the second numbers are the associated loss in million dollars (see Table 3.3). Two secondary information datasets are assumed for the high temperature event scenario. The first dataset is $[160,0.4]$ where 160 ºC is the set-point for the high temperature alarm, and it is assumed that, at this temperature, 0.4 millions of US dollars in quality loss occurs. The second dataset, $[180,4]$, is associated with the set point for the high–high temperature alarm where the emergency shutdown system (ESD) is expected to operate. These primary and secondary datasets are used to calculate the shape parameter using Equation (3.3). The shape parameter for the high temperature event loss function is obtained as $\alpha_{HTE}=2.62$. Similarly, considering $[150,0]$ and $[100,4]$ as the primary information, and $[140,0.4]$ as the secondary information (associated with the set

point for a low alarm), the shape parameter for the low temperature event loss function is obtained as $\alpha_{LTE} = 3.91$.

Figure 3.5 shows the overall loss function for the CSTR determined using Equation (3.9). The curve to the left of the target is the loss function for the low temperature event, and the curve on the right side of the target is the loss function for the high temperature event. As the reactor temperature deviates more from the target operating temperature (150 ºC) to either side and upon failure of successive layers of protection, the system experiences increasing losses.

If enough information were available, the actual system loss might have a shape like the dashed line in Figure 3.5. However, as it is not practical to obtain the actual loss behaviour, IBLF can be used as a tool to model the estimated system loss due to process deviations. The table next to Figure 3.5 represents the possible outcome scenarios, corresponding layers of protection (LOP) involved, and applicable losses at different stages of the $IBLF_{HTE}$. In Figure 3.5, BPCS denotes the basic process control system, OP is the operator, ESD refers to the emergency shutdown system, and HA and HHA denote the high temperature alarm and the high–high temperature alarm, respectively. Considering the high temperature event scenario in the CSTR, i.e. right side curve in Figure 3.5, the following outcome states are identified:

- C2-NM: Near-miss. Upon deviation of the reactor temperature from the target temperature and up to 160 ºC the BPCS is expected to operate and bring the system to normal operation. With an increase in the deviation of the reactor temperature, the undesired reaction accelerates, along with the loss of quality. In Figure 3.5, the

IBLF up to 160 ºC and the horizontal line after that represents loss associated with C2-NM.

- C2–C3: If the reactor temperature exceeds the set-point of the high alarm, the operator is expected to take corrective action. If the operator is successful, this situation is referred to as a near-miss. If the operator fails to detect and diagnose the temperature deviation, the reactor temperature may approach the set point for the high–high temperature alarm, 180 ºC, where activation of the high–high alarm and ESD system will result in process shut-down (mishap).
- C3-SD: Process shut-down. When the reactor temperature reaches 180 ºC, successful operation of the high–high temperature alarm followed by successful activation of the automatic ESD system or, if the automatic ESD is not activated, operator action to manually activate the ESD system, will result in process shut-down (mishap).
- C3–C4: Failure of the ESD system will cause the reactor to experience a runaway reaction which ultimately might cause reactor failure. The higher the reactor temperature at which the reactor fails, the more extensive the resulting damage area and the higher the resulting losses. Based on the effectiveness of physical protection and mitigation systems in place, either an incident or an accident will occur.
- C4-RA/QM: Runaway/Quality Meltdown. It is considered that at 220 ºC the reactor failure occurs in a catastrophic manner. Given that all protection layers have failed to protect the reactor or mitigate the consequences, the most severe losses are considered for this situation (disaster).

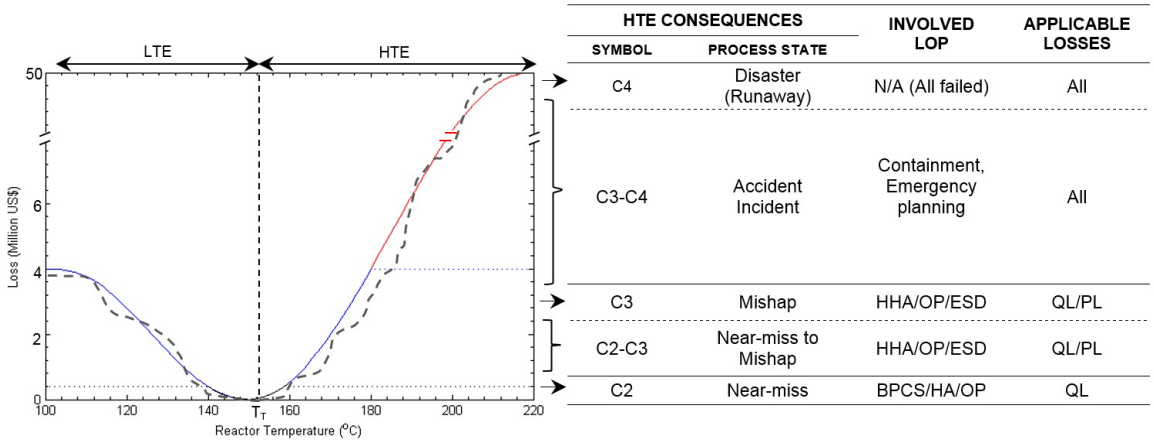


Figure 3.5. Overall loss function for the high temperature event and the low temperature event in the CSTR

3.4.3. Scenario Probability Analysis for the CSTR

The event tree analysis is used for scenario probability estimation. As discussed in Section 3.3.4, the adjusted demand rate and the probability of failure on demand of each safety barrier are used to calculate the failure frequency of safety barriers. Equation (3.6) is used to determine the demand rate adjustment factor. For illustration purposes, the shape parameter in Equation (3.6) is considered as $\alpha_{DRAF} = 3.81$. Equation (3.5) is then used to obtain the adjusted demand rate where the process deviation frequency is assumed as 1 event per month. Table 3.4 shows the considered probability of failure on demand values for the CSTR safety barriers adopted from standard handbooks, e.g. Center for Chemical Process Safety [31] and using expert judgement.

Table 3.4. Probability of failure on demand values for different safety barriers in the CSTR

SB_i	Safety Barrier	PFD (Event/year)
SB_1	Control system	0.01
SB_2	High temperature alarm	0.05
SB_3	Operator action	0.2
SB_4	Hi-Hi temperature alarm	0.05
SB_5	Automatic ESD	0.01
SB_6	Manual ESD	0.4

After the calculation of the process deviation frequency and the probability of failure on demand of the safety barriers, the failure frequency of each safety barrier is determined using Equation (3.4). Frequencies of process end states are calculated based on Equation (3.7). For instance, frequency of occurrence of process shutdown due to high temperature event, $F(C_{3-SD})$, is calculated as

$$F(C_{3_1-SD}) = IBLF_{SB1} \times (1 - IBLF_{SB2}) \times IBLF_{SB3} \times (1 - IBLF_{SB4}) \times (1 - IBLF_{SB5}). \quad (3.10)$$

Similarly, $F(C_{3_2-SD})$ to $F(C_{3_6-SD})$ are also calculated to estimate $F(C_{3-SD}) = \sum_{i=1}^{6} F(C_{3_i})$.

The frequencies of near-miss C_{2-NM} and runaway ($C_{4-RA/QM}$) for the high temperature event and probabil- ity of shutdown for the low temperature event are calculated using Equation (3.7), and the results are plotted in Figure 3.6.

As can be seen in Figure 3.6, any deviation of reactor temperature to either side of the target increases the frequency of a near-miss happening, resulting in an undesired reaction inside the reactor and loss of quality of product. If the control system fails to return the reactor temperature to the normal operating value, followed by a failure of the high-temperature alarm and operator intervention, the ESD system is designed to shut down the reactor. This situation is shown in Figure 3.6 with a dashed and dotted line, where the

frequency of process shutdown occurrence starts to increase beyond 160 ºC and below 140 ºC, the set points for high- temperature and low-temperature alarms, respectively. For the high temperature scenario, if the ESD system fails to shut down the process, the undesired exothermic reaction accelerates and the probability of runaway increases.

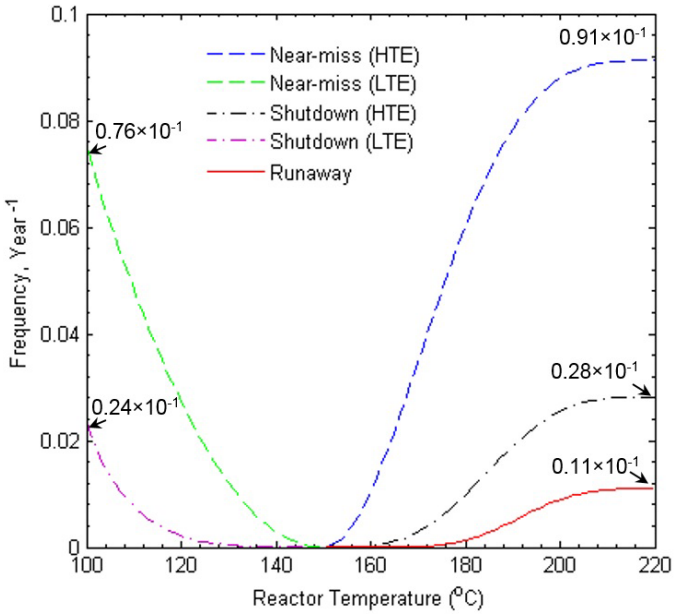


Figure 3.6. Frequency of occurrence of process end states vs. reactor temperature in the CSTR

3.4.4. Estimated Risk for the CSTR

Figure 3.7 illustrates the risk of reactor temperature deviation for both high and low temperature scenarios determined from Equation (3.8). Three curves on the right side of the reactor target operating temperature (150 ºC) represent the risk of near-miss, process shutdown, and runaway conditions due to the temperature deviations. From 150 ºC to 165 ºC (5 ºC above the high temperature alarm set point) the risks of process shutdown and runaway are negligible and the risk of quality loss is the only considerable risk. Although

the reactor does not experience unsafe situations while the temperature deviations are within the safe operating window, occurrence of such deviations imposes quality loss to the system due to production of off-specification products, material wastage, associated internal and external quality costs and so on.

If the control and safety systems fail to bring the deviated reactor temperature to its target, the reactor will experience unsafe conditions. Risk of process shutdown is shown in Figure 3.7 where the temperature starts to increase beyond 165 ºC. At 180 ºC, the automatic and/or manual ESD systems are supposed to shut down the reactor. If these protection layers also fail, the reactor will experience a runaway condition. As can be seen in Figure 3.7, runaway risk increases suddenly after the temperature exceeds the ESD set point (180 ºC). Above 200 ºC, the risk of runaway will be the dominant risk with a very sharp increased rate and the risk of process shutdown becomes constant at about 0.12 million dollars per year.

The two curves to the left of the reactor target in Figure 3.7 show the risk of near-miss (quality loss) and process shutdown due to a low temperature scenario. As shown in Figure 3.7, at low temperatures, the system experiences quality loss. If the control system fails, at temperatures below the set point of the low-alarm (140 ºC), the operator is expected to take corrective action; otherwise, the reaction stops when the temperature falls below 100 ºC. The dashed red line illustrates the total risk of the reactor, which is determined by taking the maximum risk of all potential scenarios for each given temperature range.

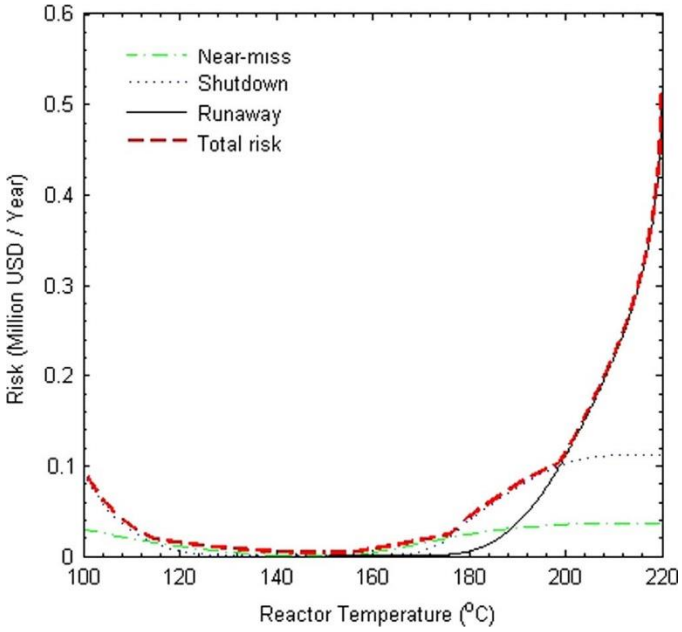


Figure 3.7. CSTR risk due to reactor temperature deviation

As shown in this case study, it is possible to analyze the effects of process deviations on both the probability and consequences of process end-states using the proposed methodology. By providing a clear representation of how the risk profile dynamically changes with process deviations, the proposed methodology facilitates monitoring of safety and quality performance in an integrated way. This will allow the overall evaluation of performance improvement strategies.

3.5. Conclusions

A loss function-based consequence evaluation method is proposed. The concept of the demand rate adjustment factor is introduced to incorporate the effect of process deviations on the demand rate of safety barriers. The adjusted demand rate is then combined with the

probability of failure on demand to determine the updated probability of failure of the safety barriers at a current process state. The proposed methodology provides risk profiles associated with different process end states. This enables integrated assessment and continuous monitoring of process safety and quality performance. In order to achieve the highest levels of safety and quality, the improvement measures need to be incorporated into day-to-day activities. Instead of relying on the safety level considered at the design stage, the estimated risk profile will enable operators to make informed decisions based on the real-time operational risk. This can also be used as a criterion to annunciate, prioritize, and analyse warnings generated as a result of process deviations. However, in practical applications a typical process system possesses multiple key characteristics related to product quality and process safety. Developing a generalized multivariate loss function-based consequence analysis methodology for process systems is a subject for future research.

3.6. References

[1] Khan F, Abunada H, John D, Benmosbah T. Development of risk-based process safety indicators. Process Saf Prog 2010;29:133–43.

[2] CCPS. Guidelines for Risk Based Process Safety. Hoboken, New Jersey: Center for Chemical Process Safety. John Wiley & Sons, Inc.; 2007.

[3] García Herrero S, Mariscal Saldaña MA, Manzanedo del Campo MA, Ritzel DO. From the traditional concept of safety management to safety integrated with quality. J Safety Res 2002;33:1–20.

[4] Hashemi SJ, Ahmed S, Khan F. Loss functions and their applications in process safety assessment. Process Saf Prog 2014;33:285–91. doi:10.1002/prs.11659.

[5] Taguchi G. Quality Engineering In Production Systems. First. New York: McGraw-Hill; 1989.

[6] Dumas R. Safety and quality: The human dimension. Prof Saf 1987;32:11–4.

[7] Krause TR. Safety and quality: two sides of the same coin. Qual Prog 1994;October:51–5.

[8] Adams EE. Total Quality Safety Management. Des Plaines, IL: American Society of Safety Engineers; 1995.

[9] Khan FI, Iqbal A, Ramesh N, Abbasi S a. SCAP: a new methodology for safety management based on feedback from credible accident-probabilistic fault tree analysis system. J Hazard Mater 2001;87:23–56.

[10] Meel A, Seider WD. Real-time risk analysis of safety systems. Comput Chem Eng 2008;32:827–40. doi:10.1016/j.compchemeng.2007.03.006.

[11] Kalantarnia M, Khan F, Hawboldt K. Dynamic risk assessment using failure assessment and Bayesian theory. J Loss Prev Process Ind 2009;22:600–6. doi:10.1016/j.jlp.2009.04.006.

[12] Khakzad N, Khan F, Amyotte P. Dynamic safety analysis of process systems by mapping bow-tie into Bayesian network. Process Saf Environ Prot 2013;91:46–53. doi:10.1016/j.psep.2012.01.005.

[13] Khan F, Abbasi S. A criterion for developing credible accident scenarios for risk assessment. J Loss Prev Process Ind 2002;15:467–75. doi:10.1016/S0950-4230(02)00050-5.

[14] Rathnayaka S, Khan F, Amyotte P. SHIPP methodology: Predictive accident modeling approach. Part I: Methodology and model description. Process Saf Environ Prot 2011;89:151–64. doi:10.1016/j.psep.2011.01.002.

[15] Marsh Risk Consulting. The 100 Largest Losses. London: Marsh Global Energy Risk Engineering; 2011.

[16] Khan F, Amyotte PR. I2SI: A comprehensive quantitative tool for inherent safety and cost evaluation. J Loss Prev Process Ind 2005;18:310–26. doi:10.1016/j.jlp.2005.06.022.

[17] API. Recommended Practice 581: Risk-Based Inspection Technology. 2nd ed. Washington: American Petroleum Institute; 2008.

[18] Pun K-F, Hui I-K. Integrating the safety dimension into quality management systems: A process model. Total Qual Manag 2002;13:373–91.

[19] Pariyani A, Seider WD, Oktem UG, Soroush M. Dynamic Risk Analysis Using Alarm Databases to Improve Process Safety and Product Quality: Part I - Data Compaction. AIChE J 2012;58:812–25. doi:10.1002/aic.12643.

[20] Arunraj NS, Maiti J. A methodology for overall consequence modeling in chemical industry. J Hazard Mater 2009;169:556–74. doi:10.1016/j.jhazmat.2009.03.133.

[21] CCPS. Guidelines for Consequence Analysis of Chemical Releases. New York: Center for Chemical Process Safety of the American Institute of Chemical Engineers; 1999.

[22] British Standards Institute. BS 6143: Guide to the Economics of Quality, Part 2. Prevention, Appraisal, Failure Model. London: BSI; 1990.

[23] Margavio GW, Fink RL, Margavio TM. Quality Improvement Using Capital Budgeting and Taguchi's Function. Int J Qual Reliab Manag 2005;11:10–20.

[24] Leung BPK, Spiring FA. Some Properties of the Family of Inverted Probability Loss Functions. Qual Technol Quant Manag 2004;1:125–47.

[25] Leung BPK, Spiring FA. The inverted beta loss function: properties and applications. IIE Trans 2002;34:1101–9.

[26] ISA-TR84.00.02 - Part 2. Safety Instrumented Functions (SIF)-Safety Integrity Level (SIL) Evaluation Techniques Part 2: Determining the SIL of a SIF via Simplified Equations. North Carolina: ISA - The Instrumentation, Systems, and Automation Society; 2002.

[27] Musharraf M, Hassan J, Khan F, Veitch B, MacKinnon S, Imtiaz S. Human reliability assessment during offshore emergency conditions. Saf Sci 2013;59:19–27. doi:10.1016/j.ssci.2013.04.001.

[28] Dowell III AM. Layer of protection analysis for determining safety integrity level. ISA Trans

1998;37:155–65.

[29] Willis MJ. Some Conventional Process Control Schemes. Lect Notes Distrib Process Control 2 Univ Newcastle 1999.

[30] Meel A, Seider WD. Plant-specific dynamic failure assessment using Bayesian theory. Chem Eng Sci 2006;61:7036–56. doi:10.1016/j.ces.2006.07.007.

[31] CCPS. Guidelines for Chemical Process Quantitative Risk Analysis. Second. New York: Center for Chemical Process Safety of the American Institute of Chemical Engineers; 2000.

4. CORRELATION AND DEPENDENCY IN MULTIVARIATE PROCESS RISK ASSESSMENT[3]

Preface

A version of this manuscript has been published in the proceedings of the IFAC SAFEPROCESS 2015. I am the primary author of this paper. Along with the co-authors, Faisal Khan and Salim Ahmed, I formulated the problem and carried out most of the literature review, data collection and analysis. I prepared the first draft of the manuscript and subsequently revised the manuscript based on the co-authors' feedback and also the peer review process. The co-author Faisal Khan helped in developing the concepts/models and their testing, reviewed and corrected the models and results, and contributed in preparing, reviewing and revising the manuscript. The co-author Salim Ahmed contributed through support in the development, testing and improvement of the model. Salim Ahmed also assisted in reviewing and revising the manuscript.

Abstract

Process safety and risk assessment are often multidimensional and hence require the joint modeling of several potentially correlated random variables. Any effort to address the correlation among the input variables is important and could improve the accuracy in practical applications of risk assessment models. This paper discusses the problems with

[3] Hashemi et al. IFAC SAFEPROCESS 2015: 9th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes., vol. 48, Paris; 2015, p. 1339–44.

correlated variables used in risk assessment and presents a copula-based technique to model dependency among variables to improve uncertainty analysis. Using the copula approach, capturing the dependence structure among different risk factors and estimating the univariate risk marginals can be separated. This advantage simplifies the overall risk estimation for systems with multiple dependent risk sources. The advantage of the copula-based framework for generalization over the traditional correlation analysis technique is demonstrated using a case study. Methods are also presented for copula selection and estimation of the copula parameters.

Keywords: Correlation coefficients; dependence; copula function; copula estimation; safety analysis**.**

4.1. Introduction

In recent years there has been an increased attention in the integrated management of risk in process industries. It is no longer the best practice to consider each risk factor in isolation. Correlations among the factors and their potential synergy to cause catastrophic losses need attention. Thus, understanding the joint distribution of all risk sources is of paramount importance in process industries. Choosing and estimating a useful form for the marginal distribution of each variable in its domain is often a straightforward task. In contrast, other than the normal and *t*-distributions, univariate distributions usually do not have a convenient multivariate generalization [1]. Moreover, for these two families, the marginal distributions are also normal or *t*-distributed, respectively. This restriction limits their application to practical situations. Indeed, modelling and estimation of flexible

(skewed, multi-modal, heavy tailed) high-dimensional distributions is still an existing challenge. Developing a framework which allows specifying the marginal distributions irrespective of the dependence structure is a potential approach to address this challenge.

Risk practitioners need to deal with complex process systems with multiple correlated variables. However, the available mathematical tools to analyse, extract and make use of their correlation information are limited. The best known tool has been the linear correlation coefficient. Linear correlation, or Pearson correlation, is a global measure that attempts to summarize the dependence between two variables using a single number. It cannot be expected to adequately summarize complex dependencies into a single number [2]. As shown later in this paper, two datasets with different dependence patterns can have the same correlation coefficient.

Copula functions [3] offer a general framework for constructing multivariate distributions using any univariate marginals and a copula function C that links these marginals. The copula approach is important from a modelling perspective as it provides a tool to separate the choice of the marginals and that of the dependence structure which is expressed in C [1]. In practice, this advantage often lead to significant improvement in the analysis of multivariate systems. Accordingly, there has been an increased interest in copulas in oil and gas industries, with applications ranging from process safety assessment [4] to the estimation of oil well drilling duration [5].

4.2. The Correlation Challenge

The word “correlation” has been frequently used (or misused) as an over-arching term to describe all sorts of dependence between two random variables [6]. However, correlation is only one particular measure of stochastic dependency among many. It is the canonical measure in the world of multivariate normal distributions, and more generally for spherical and elliptical distributions [7]. However, several researches in process loss modelling and risk assessment shows that the distributions of the real world are seldom in this class [4]. Figure 4.1 represents the motivation of this study, where 1000 bivariate realization from two different risk models for (R_1, R_2) are shown. In both models, R_1 and R_2 have identical log-normal and Weibull marginal distributions, respectively, and the linear correlation between them is 0.75. However, it is clear that the dependence between R_1 and R_2 in the two models is qualitatively different. Moreover, if we consider the random variables to represent risk classes, the second model is more dangerous from the point of view of a process risk analyst, since extreme losses generally occur together. We will return to this example later in the paper. For the time-being, as pointed out by [7], it should be noted that the dependence in the two models cannot be distinguished on the grounds of correlation alone.

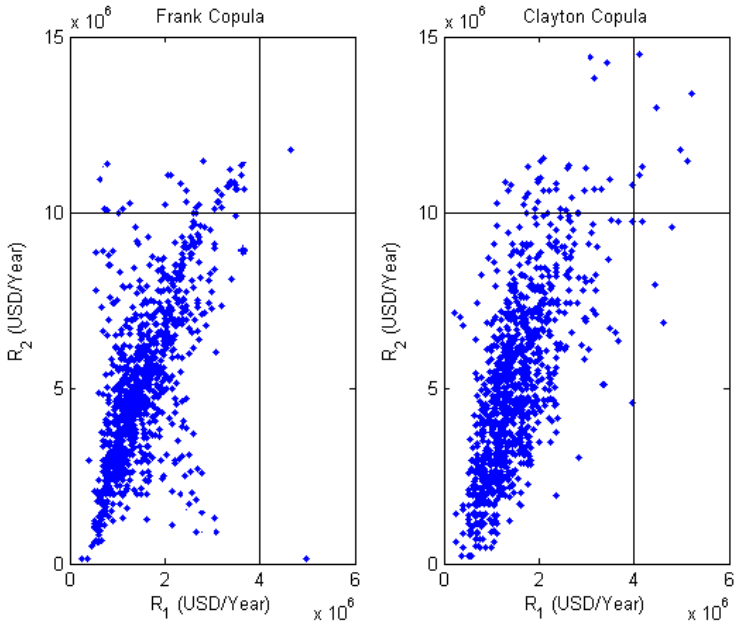


Figure 4.1. 1000 bivariate realization from two distributions with identical marginal distributions and the linear correlation 0.75, but different dependence structure

The main objective of this paper is to describe the importance of correlation and dependency in the context of multivariate risk assessment of process industries. The concept of copulas is discussed in Section 4.3 as an effective tool to capture and demonstrate the dependency. Section 4.4 collects and clarifies the essential ideas of dependence, linear correlation (and its shortcomings), and rank correlation, which have not been addressed properly in related process industries literature. Section 4.5 discusses the importance of considering dependence among both frequencies and consequences of abnormal events when analysing the aggregate risk of a process. Section 4.6 is another contribution of this study that addresses two challenges in application of copula-based approaches: the copula parameter estimation and copula model selection. Using a simple case study, the application of nonparametric and semi-parametric methods to estimate

copula parameters are presented in Section 4.6 and the difference between these two estimates is then used to select the best fitting copula.

4.3. Copulas

Copulas are used to describe the joint distribution of dependent random variables. With copula modelling, the marginal distributions from different families can be combined [8]. This is the main advantage of copulas compared with alternative methods, such as the use of multivariate distributions, to construct dependencies.

Let $\boldsymbol{R} = R_1, \ldots, R_d$ be a random vector with continuous marginal cumulative distribution functions (CDF) $F_1, \ldots, F_d$. Based on Sklar's theorem, the CDF H of $\boldsymbol{R}$ can be represented as:

$$H(R_1,...,R_2) = C\{F_1(R_1),...,F_d(R_d)\} \tag{4.1}$$

for all real $R_1, \ldots, R_d$ in terms of a unique function C: $[0, 1]^d \rightarrow [0, 1]$, called a copula [3]. To provide a formal definition for copulas, let $U_1, \ldots, U_d$ be real random variables marginally uniformly distributed on [0, 1]. A copula function is a joint distribution function defined as

$$C(u_1,...,u_d) = P(U_1 \le u_1,...,U_d \le u_d). \tag{4.2}$$

The representation in Equation (4.1) is very fundamental for a copula application in risk assessment; it indicates that given any marginal risk distribution ($R_1, \ldots, R_d$) and a copula function C, Equation (4.1) can be used to obtain the joint risk distribution function. The main advantage provided to the process risk analysts by this representation is that the

selection of an appropriate model for the dependence between risk sources, represented by *C*, can then proceed independently from the choice of the marginal risk distributions.

Nelsen [3] reviewed different copula families and discussed methods of constructing copulas. For instance, Table 4.1 shows three common families of Archimedean copulas and their parameter space.

Table 4.1. Three common families of Archimedean copulas and an expression for the population value of Kendall's τ

Family	***C*(*u*, *v*)**	$\delta \in$	τ
Clayton	$\left(u^{-\delta}+v^{-\delta}-1\right)^{-1/\delta}$	$\delta \le 1$	$\delta/(\delta+2)$
Gumbel	$e^{\left\{-\left[(-\ln u)^{\delta}+(-\ln v)^{\delta}\right]\right\}^{1/\delta}}$	$\delta \ge 1$	$1-1/\delta$
Frank	$\left(-\frac{1}{\delta}\right)\times\log\left\{1+\frac{\left[e^{-\delta u}-1\right]\cdot\left[e^{-\delta v}-1\right]}{e^{-\delta}-1}\right\}$	$\delta \in \mathbb{R}$	$1-4/\delta+4D_1(\delta)/\delta$

Note: Here, $D_1(\delta)=\int_0^{\delta}(x/\delta)/(e^{x}-1)dx$ is the first Debye function.

Copula is sometimes referred as a "dependency function" since it contains all of the dependence information between random variables [6]. For instance, using the Frank copula, all the information about the dependence between the two random variables is in the parameter δ, whose value can be interpreted in terms of a coefficient like Kendall's τ rank correlation. Different measures of dependence (correlation coefficients) are discussed in the next section.

4.4. Dependence Measures

Two kinds of dependence measures are briefly discussed in this section; the traditional Pearson linear correlation, and the rank correlation. The latter measure is based on a copula

function, however, both of the measures calculate scalar measurement for a pair of random variables. Nevertheless, the specification and the nature behind each of them varies. In the following section, at first, the definition of the linear correlation and its shortcomings when applied in non-elliptical models are provided. Then, the rank correlation and its representative, the Kendall's τ rank correlation, are described.

4.4.1. Linear Correlation

The word "correlation" is used in this work only in its technical sense of linear correlation or Pearson's correlation, denoted by ρ. Let X and Y represent two random variables with non-zero finite variances. The linear correlation coefficient for (X, Y) is:

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}, \tag{4.3}$$

where *Cov* and *Var* are the covariance and variance operators, respectively. This well-known measure of correlation has the following properties:

- $\rho(X,Y) \in \langle -1,1 \rangle$,
- $\rho(X,Y) = |1|$ means perfect correlation, positive or negative,
- $\rho(X,Y) = 0$ indicates no correlation between the random variables.

However, it needs to be stated that if two random variables are not correlated, it does not mean that they must be independent. No correlation indicates that there is no dependency only under normality. Correlation is considered to be only one particular measure of stochastic dependence among many others [6]. Another drawback of linear correlation is its assumption of finite variances of X and Y. This could be a problem when this measure

is to be applied on heavy-tailed distributions, where the variance of random variables may be infinite [7].

Though linear correlation is a popular measure of dependence, it is often misinterpreted. As pointed out by Klaus (2012), the wide application of correlation is primarily due to its simple computation. Moreover, it is a natural scalar measure of dependence for elliptical distributions (for example for multivariate normal distribution). Nevertheless, using correlation as a measure of dependence by assuming that all multivariate distributions are elliptically distributed would produce misleading results in real life problems.

4.4.2. Rank Correlation

The copula of two random variables completely determines any dependence measures that are scale-invariant, that is, measures that remain unchanged under monotonically increasing transformations of the random variables. More generally, if φ and Ψ are two increasing transformations with inverses φ^{-1} and Ψ^{-1}, the copula of the pair (Z, T) with $Z = \varphi(X)$ and $T = \Psi(Y)$ is the same as that of (X, Y); for a proof see [8]. Expressed in different terms, the construction of the multivariate distribution in Equation (4.1) implies that the copula function C is invariant under monotonically increasing transformations of its margins. Therefore, scale-invariant dependence measures can be expressed in terms of a copula-based measure of dependence, the rank correlation. Unlike linear correlation, rank correlation does not depend on marginal distribution but only on copula [9].

The most widely used rank correlation, also known as concordance measure, is Kendall's τ. Informally, a pair of random variables are concordant if large (or small) values of one

tend to be associated with large (or small) values of the other [3]. The Kendall's τ can be defined by introducing a concordance function between two continuous random vectors (X, Y) and (X', Y') with possibly different joint distributions, but with common marginal distributions.

For a bivariate random vector (X_1, X_2) with copula C, Kendall's τ is interpreted as the difference between the probability of concordance and dis-concordance of two independent and identically distributed observations [3]:

$$\tau = 4\int_0^1\int_0^1 C(u_1,u_2)dC(u_1,u_2) - 1. \tag{4.4}$$

Note that τ is a symmetric dependency measure and takes values $\tau \in \langle -1,1 \rangle$, where -1 signals a perfect negative correlation, 1 displays a perfect positive correlation and 0 shows no correlation. However, similar to linear correlation, this does not indicate independency [6].

Spearman's ρ is another rank correlation which is proportional to the difference between the probability of concordance and dis-concordance of two vectors. There are also other dependence measures based on copulas. For example, tail dependence is a very important measure when studying the dependence between extreme events. These dependence measures are not within the scope of this work; an interested reader may refer to [3].

4.4.3 Choosing Dependence Measures

There are numerous guidelines on when to use each correlation coefficient, though their suggestions sometimes contradicts each other [10]. However, based on [10], a common

practice is not to use the linear correlation coefficient for non-normal data with obvious outliers and for highly skewed distributions.

To wrap up this section, it can be concluded that while the linear correlation coefficient is still needed to parameterize the underlying bivariate normal, rank correlations are more useful in describing the dependence between random variables, because they are invariant to the choice of marginal distribution.

4.5. Dependence in Risk Assessment

Although the application of risk-based fault detection [11] and warning system design [12] approaches are increasing in the process industries literature, the effect of dependence among risk model parameters have not been studied well yet.

As an example of a system with two risk sources, let us denote R_1 and R_2 as the aggregate operational and business risks, respectively, with the following:

$$R = R_1 + R_2 = \sum_{n=1}^{N_1} L_{1n} + \sum_{m=1}^{N_2} L_{2m} \tag{4.5}$$

where R is the global aggregate risk, L_1 and L_2 are loss amounts (i.e. severity of events) and N_1 and N_2 are the frequencies of each loss. When incorporating the dependence structure in risk assessment, it is necessary to clarify which dependence we are talking about since each type of risk is driven by two elementary sources of randomness, i.e. frequency (or probability) and loss (or consequence). In this respect, as described in the following sections, dependence among aggregate risks may result from dependence among frequencies or among losses or between both.

4.5.1. Frequency Dependence

Intuitively, frequency dependence means that, historically, the number of (for example) mishaps resulted in both operational and business losses is high when the number of (for example) near-misses resulted in process variations and consequent operational loss is also high. It likely happens when both frequencies N_1 and N_2 share common dependence with respect to some variables such as the performance of the control system and safety barriers. Empirically, frequency correlation could be evidenced and measured by computing the historical correlation between past frequencies of events, provided of course that data are recorded for a sufficiently long period of time [13]. Frequency dependence and its effect on aggregate risk assessment is an interesting subject for future research.

4.5.2. Loss Dependence

Mathematically, loss dependence would mean that loss amounts randomly drawn from different classes of events are not independent of one another. It may be observed when, for example, operational loss amounts due to the production of off-specification product are high (or low), then reputational and business loss amounts are also high (or low). Empirically, the correlation among different loss classes can be identified by monitoring the business performance and loss amounts of a given process over time. The concept of loss dependence and the application of copula functions for aggregation of dependent loss random variables is studied in an earlier work by the authors [4].

4.6. Case Study

A detailed study on how and to what extent frequency and/or loss dependence impact the aggregate risk is a subject for a future study. However, this section aims to illustrate the significance of observing the dependence among random variables when conducting risk analysis. For this purpose, the motivating example in the Introduction section is used to investigate the effect of dependence between operational and business risk classes of a hypothetical distillation process when estimating the global aggregate risk.

From the operational history of the distillation process, the operational risk (R_1) due to the deviation of the distillation column key process variables from their target values is estimated to follow a log-normal distribution with mean μ = 0.35 million US dollars (USD) per year and standard deviation σ = 0.42 million USD/year. Analysing the process business performance within the past 12 months, the business risk (R_2) of the process due to interrupted production is considered to be Weibull distributed with shape parameter β_w = 5.79 and scale parameter θ_w = 2.16 million USD/year. The focus of this case study is on modelling the dependence between the two risk classes. Sensitivity analysis investigations are provided in Sections 7.6.4 and 8.4.5 to evaluate the impact of uncertainty in data on estimated correlation parameter.

4.6.1. Assessment of Dependence

Before a copula model for the pair (R_1, R_2) is determined, the scatter plot of the data was used to check for the presence of dependence. The scatter plot of (R_1, R_2) pairs shown in Figure 4.2 suggests the presence of positive association between operational and business

risks, as might be expected. Figure 4.2 has histograms alongside a scatter plot to show both the marginal distributions of R_1, R_2, and the dependence. To quantify the degree of dependence in the pair (R_1, R_2), sample value of Kendall's τ and Pearson ρ are determined in Table 4.2 using the MATLAB® [14] software `corr` function. According to a simple comparison in Table 4.2, as discussed in Section 4.4.2, transformation of variables does not affect the rank-based Kendall's τ correlation coefficient.

Table 4.2. Comparison between the Pearson ρ and Kendall's τ values

Dependence Measure	$R_1 \times R_2$	Log (R_1) × Log (R_2)
Pearson ρ	0.75	0.70
Kendall's τ	0.56	0.56

In order to model the dependence between the operational risk (R_1) and the business risk (R_2) of the distillation process, only the class of Archimedean copulas are considered in this work for simplicity. The simple closed functional forms of Archimedean copulas along with their desired properties made them suitable for variety of applications [3,5]. Three widely used Archimedean models used in the literature are: Gumbel's family with upper tail dependence; Clayton's family with lower tail dependence; and Frank's family that has reflection symmetry.

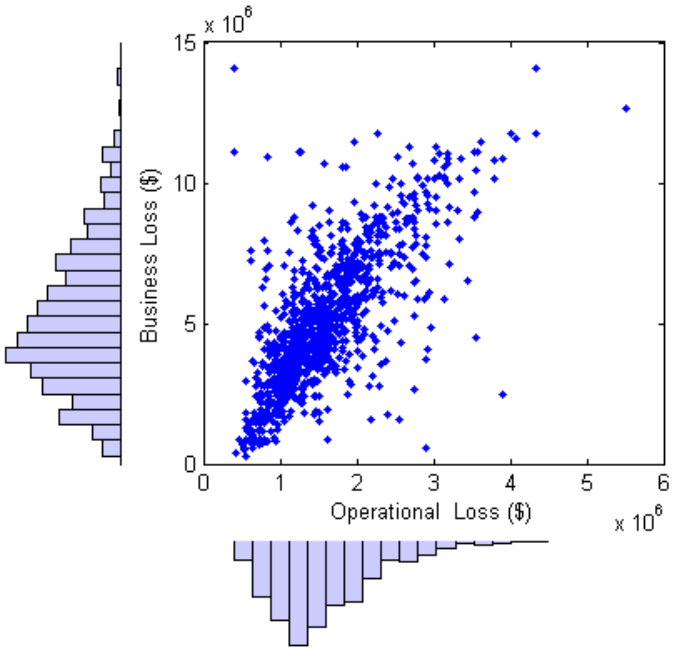


Figure 4.2. Business risk and operational risk distributions for the distillation process data

4.6.2. Copula Estimation

Next step to model the joint risk density is to estimate the copula parameters. Among different techniques to estimate and select copulas, a nonparametric estimation of the copula [15] and a semi parametric estimation based on the method of maximum pseudolikelihood [16] are presented in this work. After obtaining both estimates, the results are compared to select the best fitting copula.

Genest and MacKay [17] presented that for parametric copulas, such as Archimedean models, one can estimate the copula parameter from the closed form relationship between the copula parameter and Kendall's τ. As discussed in Section 4.4.2, Kendall's τ is the difference between the probability of concordance and the probability of discordance of independent pairs of realizations of a joint distribution. Therefore, the sample version of

Kendall's τ is given by $\hat{\tau} = (c-d)/(c+d)$, where c denotes the number of concordant pairs and d the number of discordant pairs in the sample [18]. Table 4.1 gives an expression for τ for the three most common Archimedean models. An advantage of this approach is that the marginal distributions do not need to be specified, however, this method is only applicable to the one parameter copulas [18]. For the estimated Kendall's τ value of 0.56, the nonparametric estimates of each copula family parameter are obtained as $\hat{\delta}^{NP}_{Clayton} = 2.54$, $\hat{\delta}^{NP}_{Gumbel} = 2.27$, and $\hat{\delta}^{NP}_{Frank} = 6.95$.

The other copula estimation procedure is the canonical maximum likelihood (CML), or the semi-parametric estimation, method proposed by [16]. In this two-step estimation procedure, at first the empirical distribution functions of the series of interest for each $1 \leq i \leq p$ are determined as

$$\hat{F}_{in}(x) = \frac{1}{n+1}\sum_{j=1}^{n}\mathbf{1}(X_{ij} \leq x) \tag{4.6}$$

where $\mathbf{1}(X_{ij} \leq x)$ is the indicator function. Then, the semi-parametric estimate of the copula parameter δ is the value that maximize the log-likelihood function of the copula density using the transformed variables given by

$$\ell(\delta) = \sum_{k=1}^{n}\log\left\{c_{\delta}\left[F_{1n}(X_{1k}),...,F_{pn}(X_{pk})\right]\right\} \tag{4.7}$$

where c_{δ} is the copula density. Although the numerical work involved with this method may make it less attractive, however, the CML method is generally applicable for more copula models. Moreover, application of the CML method does not require the marginal distribution to be known and also does not require the dependence parameter to be real.

Given the copula families in Table 4.1, the copula parameters for the current case study were obtained using the presented CML methodology for sample size of n = 10,000 as $\hat{\delta}^{SP}_{Clayton} = 2.45$, $\hat{\delta}^{SP}_{Gumbel} = 4.38$, and $\hat{\delta}^{SP}_{Frank} = 13.15$.

4.6.3. Copula Selection

Given that there exists only three competing copulas for this case study, the absolute value of the distance between the nonparametric and semi-parametric copula parameter estimates, $\Delta_{\hat{\delta}} = \left|\hat{\delta}^{NP}_{C} - \hat{\delta}^{SP}_{C}\right|$, is used as a criterion in this case study to select the best copula. When the $\Delta_{\hat{\delta}}$ value is small, one has an indication of a reasonable fit. In this case study, the $\Delta_{\hat{\delta}}$ values for different copulas are determined as: $\Delta^{Clayton}_{\hat{\delta}}$ = 0.09, $\Delta^{Gumbel}_{\hat{\delta}}$ = 2.11, and $\Delta^{Frank}_{\hat{\delta}}$ = 6.20. Thus, the Clayton with the smallest $\Delta_{\hat{\delta}}$ is selected as the best fitting copula for this case study. This result is not surprising as the scatter plot of the risk data observations in Figure 4.2 matches the lower tail dependence property of Clayton copula.

4.6.4. Overall Risk Estimation

Having estimated the copula parameters, each copula is used to simulate the joint density and marginal distributions of R_1 and R_2. For instance, the joint density of the overall risk simulated from Frank and Clayton copulas are shown in Figure 4.1. The mean (μ_R) and standard deviation (σ_R) of the aggregate risk (summation of R_1 and R_2) simulated from each copula are then determined as: $\mu_R^{Frank} = 7.176$ and $\sigma_R^{Frank} = 3.232$; $\mu_R^{Clayton} = 6.431$ and $\sigma_R^{Clayton} = 2.991$; $\mu_R^{Gumbel} = 6.705$ and $\sigma_R^{Gumbel} = 3.189$ (all numbers are in million USD/year). From Figure 4.1 and

the estimated aggregate risk values, it can be clearly seen that miss-estimation of the dependence structure may have a significant effect on the estimated aggregate risk of a process.

In this case study, it is assumed that the joint risk distribution could be well represented by of the three Archimedean models. Obviously, consideration of all potential copula models is required to ensure selection of the best fitting model. This will require a more formal and adequate copula selection method. Application of the information theory to copula selection is a subject for an ongoing research to address this challenge.

4.7. Conclusions

Deviation of process characteristics along with the failure of control systems and safety barriers cause undesired process events. This paper discusses the importance of considering correlation and dependency among frequencies and loss severities of such events for the purpose of integrated risk management. Both linear and rank correlation coefficients are introduced as the most commonly used measures of dependence in data analysis. These correlation coefficients can be represented as differently weighted averages of the same concordance indicators. The copula-based Kendall's correlation coefficient is preferable for risk assessment purposes due to usually a non-normal distribution of model parameters with non-linear dependence structures.

This paper also demonstrates the flexibility and strength of copula-based approaches in modelling the dependence among random variables. The practical application of copula-based risk aggregation and the importance of considering the correlation and dependency

in risk assessment of process industries are highlighted using a case study. The findings from the case study highlight the fact that selection of wrong dependence structure and/or wrong estimation of the copula parameter(s) could result in over or under-estimation of the overall risk.

The integration of proposed copula-based dependence modelling with existing risk-based fault detection and warning generation methods can improve the accuracy of the estimated risk by decreasing the uncertainty involved in the exiting risk models due to the assumption of independent model parameters. The copula Bayesian networks, recently introduced in actuarial studies [1], is an interesting topic for future study to capture both the connectivity and dependency in probability estimation models. By combing a copula Bayesian network with the copula-based loss aggregation methodology proposed in an earlier work [4], a new dynamic risk assessment model will be developed for process safety monitoring and warning generation of process industries.

4.8. References

[1] Elidan G. Copula Bayesian Networks. Adv. neural Inf. Process. Syst. (NIPS 2010), Vancouver: 2010, p. 559–67.

[2] Schirmacher D, Schirmacher E. Multivariate Dependence Modeling Using Pair-Copulas. 2008.

[3] Nelsen RB. An Introduction to Copulas. 2nd ed. New-York: Springer–Verlag; 2006.

[4] Hashemi SJ, Ahmed S, Khan FI. Loss scenario analysis and loss aggregation for process facilities. Chem Eng Sci 2015;128:119–29. doi:10.1016/j.ces.2015.01.061.

[5] Accioly R, Chiyoshi FY. Modeling dependence with copulas: a useful tool for field development decision process. J Pet Sci Eng 2004;44:83–91. doi:10.1016/j.petrol.2004.02.007.

[6] Klaus M. Multivariate Dependence Modeling using Copulas (Master's thesis). Charles University in Prague, 2012.

[7] Embrechts P, McNeil AJ, Straumann D. Correlation and dependence in risk management: properties and pitfalls. Risk Manag. value risk beyond, 2002, p. 176–223.

[8] Genest C, Favre A-C. Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask. J Hydrol Eng 2007:347–68.

[9] Yan J. Multivariate modeling with copulas and engineering applications. Springer Handb. Eng. Stat., Springer London: 2006, p. 973–90.

[10] Chok NS. Pearson's versus Spearman's and Kendall's correlation coefficients for continuous data. University of Pittsburgh, 2010.

[11] Zadakbar O, Khan F, Imtiaz S. Dynamic risk assessment and fault detection using a multivariate technique. Process Saf Prog 2013;32:365–75. doi:10.1002/prs.11609.

[12] Chang Y, Khan F, Ahmed S. A risk-based approach to design warning system for processing facilities. Process Saf Environ Prot 2011;89:310–6. doi:10.1016/j.psep.2011.06.003.

[13] Frachot A, Roncalli T, Salomon E. The Correlation Problem in Operational Risk. 2004.

[14] MathWorks. MATLAB Release 2013a 2013.

[15] Genest C, MacKay J. The joy of copulas: bivariate distributions with uniform marginals. Am Stat 1986;40:280–3.

[16] Genest C, Rivest L-P. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. Biometrika 1995;3:543–52.

[17] Genest C, MacKay J. The Joy of Copulas: Bivariate Distributions with Uniform Marginals. Am Stat 1986;40:280–3. doi:10.2307/2684602.

[18] Manner H. Estimation and Model Selection of Copulas with an Application to Exchange Rates. Maastricht, Netherlands: 2007.

5. LOSS SCENARIO ANALYSIS AND LOSS AGGREGATION FOR PROCESS FACILITIES[4]

Preface

A version of this manuscript has been published in the Journal of Chemical Engineering Science. I am the primary author of this paper. I formulated the problem and proposed alternate solutions. Along with the co-authors, Faisal Khan and Salim Ahmed, I developed the conceptual model and subsequently translated this to the numerical model. I carried out most of the data collection and analysis. I prepared the first draft of the manuscript and subsequently revised the manuscript based on the co-authors' feedback and also the peer review process. The co-author Faisal Khan helped in developing the concepts/models and their testing, reviewed and corrected the models and results, and contributed in preparing, reviewing and revising the manuscript. The co-author Salim Ahmed contributed through support in the development, testing and improvement of the model. Salim Ahmed also assisted in reviewing and revising the manuscript.

Abstract

This study presents an overall loss modelling methodology for process facilities. The methodology comprises loss scenario identification and aggregation of losses due to process deviations. The identification of loss scenarios and determination of the time

[4] Hashemi et al. Chemical Engineering Science 2015;128:119–29.

periods at which a process experiences each scenario are described first. Then, the application of the copula functions and their integration with the Monte Carlo (MC) approach are proposed to address the existing challenge of loss aggregation for multiple-loss scenarios. The proposed loss aggregation provides a flexible and realistic approach to construct joint multivariate distribution of the losses by considering their interdependence. The sensitivity of the model to the choice of correlation parameters is investigated. The results serve as a reminder to risk analysts about the significance of choosing an appropriate loss aggregation model for risk analysis purposes. The application of the methodology is demonstrated using a distillation column case study.

Keywords: Loss modeling; Loss aggregation; Copula function; Distillation column flooding

5.1. Introduction

Any process is subject to deterioration with time due to natural and assignable causes. As a result, the process characteristics, for example a reboiler heat duty, deviate from the specification limits and the process experiences unsafe situations that can eventually impose different losses. The losses due to abnormal conditions in process facilities can be classified into: Class 0—allowable operational loss; Class I—unallowable operational loss; Class II—business losses; and Class III—event losses. An overall loss modelling approach should identify, estimate, and aggregate all applicable significant loss elements.

The concept of identifying potential consequences and process outcomes due to abnormal events has been widely studied, usually using event tree analysis [2–6]. However, there has

been less effort to consider the time-dependent process deterioration in estimating the duration of loss scenarios. One of the contributions of this paper is to introduce a methodology to estimate the duration of loss scenarios based on the process history and safety performance. This has been integrated with the loss aggregation methodology to quantify the overall loss of the scenarios. The novelty of the proposed method is to address the time-dependent process deterioration with non-constant process mean and variance.

An important aspect of an overall loss model is the ability to aggregate the negative outcomes of scenarios with multiple losses in process facilities. This usually requires aggregation of a diverse range of loss categories with different loss distributions. More importantly, different losses are usually interrelated. However, most of the existing loss models for process facilities assume independence among these loss elements while estimating the overall loss. This simplification may cause misestimation of the overall loss. Another contribution of this paper is to address the issue of determining the aggregated loss resulting from an incident when the losses belonging to different classes are dependent. Oil spills, for example, can harm people, damage both the physical assets and the surrounding environment, and at the same time can affect the reputation of the organization with negative economic consequences. In recent years, modelling the dependencies using copulas has become popular in the actuarial, insurance, and finance literature [7,8]. Copulas may be used to construct joint multivariate distribution of losses and are rather flexible and realistic in terms of allowing a wide range of dependence structure. This paper provides a loss aggregation method based on the superposition principle. Then, aggregation methodologies based on copula functions are proposed, considering the dependence both

among the elements of a loss class and also among different loss classes. Thus, based on available information, the user can choose an appropriate aggregation method on a case-by-case basis.

The remainder of the paper has been structured as follows. Section 5.2 presents framework of the proposed overall loss modelling methodology and a loss scenario analysis procedure based on time-dependent deterioration of process characteristics. Section 5.3 reviews the existing loss aggregation approaches and provides the necessary technical background on copulas. Section 5.4 discusses the methodology and assumptions used in developing a copula-based loss aggregation method. The practical application of the methodology is illustrated in Section 5.5 using a distillation column case study. Finally, the paper is concluded with remarks on the proposed methodologies, their limitations, and potential direction for future research on the subject of using the proposed overall loss model in operational risk analysis.

5.2. Loss Scenario Modelling

The framework of the proposed overall loss modelling methodology is demonstrated in Figure 5.1. In the proposed methodology, the estimated duration of an abnormal process is required to determine the overall magnitude of time-dependent loss classes, which are operational and business losses. This section provides models to estimate time periods at which a process may experience each loss scenario. In this work, a loss scenario is defined as a description of a predicted loss situation based on process conditions and may include a single loss class or a combination of classes.

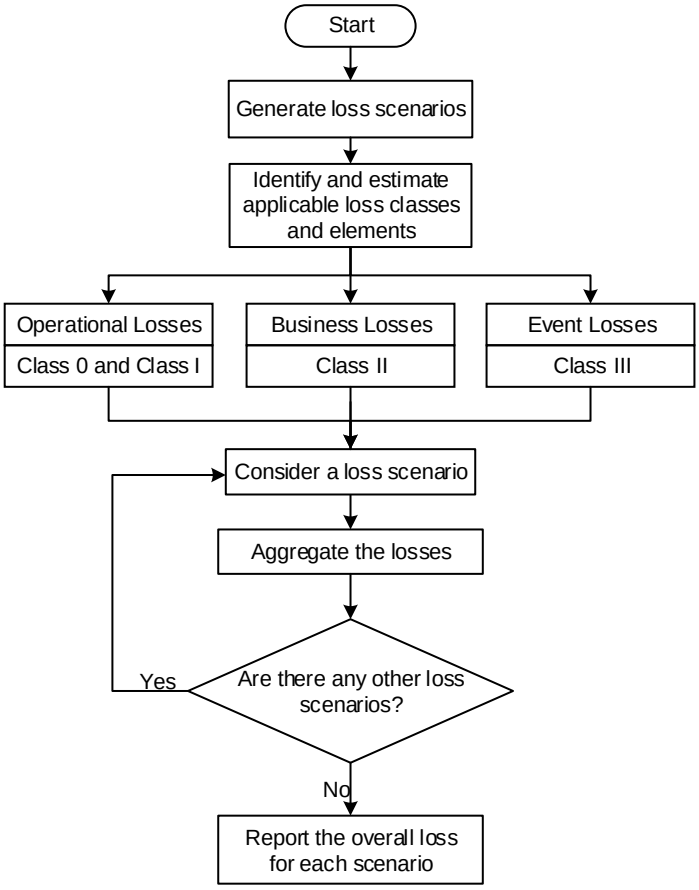


Figure 5.1. Methodology for overall loss modelling of process facilities

5.2.1. Time-Dependent Process Deterioration

Process operations are subject to deterioration with time due to chance and assignable causes. It is not possible to maintain a zero variation among materials, methods, operators, equipment, instruments, and measurements. It is process deterioration that moves the process characteristic into abnormal (out-of-control) states and increases the process variance at a random point in time [9]. These conditions impose different loss classes to the system based on the magnitude of the deviation and the performance of safety barriers. Tahera et al. [9] reviewed the issue of time-dependent process deterioration in the

manufacturing industry and proposed methods to determine quality loss considering the effect of non-constant process variance. This section considers the issue of process deterioration in chemical facilities and the resulting loss scenarios.

Consider a process system with a specific key characteristic x that is normally distributed with target mean μ_0 and an initial variance σ_0^2. It is assumed that an assignable cause occurs at a random point in time (τ_D) and changes the mean and variance in a time-dependent manner. As described in Table 5.1, according to the performance of the control system and the safety barriers, the process characteristic value could fall in four regions at a given point of time. Different process states associated with each process characteristic region and the typical actions to address the associated abnormal situations are described in Table 5.1. Table 5.2 summarizes the four loss scenarios and applicable loss elements based on the process characteristic value at time t.

The scope of this work is to develop models to estimate overall loss associated with each loss scenario, which are provided in the following section. For this purpose, it is assumed that the distribution of loss elements associated with each scenario are available.

Table 5.1. Different process characteristic regions and associated process states

Process Characteristic Region	Process State	Description	Required Action
Central Region	Normal—Type 0 (Acceptable)	Process is controlled and it is operating under a stable system of natural causes.	Process characteristic should be measured and compared with process target to ensure continued normal operation.
Warning Region	Abnormal Type 1 (Controllable)	Control system has failed to maintain the process characteristic within the specification limits and process is under the effects of assignable causes.	Raise alarm. While the process is in-service, preventive measures such as operator action are required to identify, diagnose, and eliminate the assignable cause(s) and to return the process to normal operation.
Action Region	Abnormal Type 2 (Manageable)	Preventive measures have failed to return the deviated process characteristic within the specification limits and process is under the effects of assignable causes.	Raise alarm and immediately shut down the process or relieve the pressure (if applicable). While the process is stopped, identify, diagnose, and eliminate the assignable cause(s) and restart the process.
Mitigation Region	Abnormal Type 3 (Mitigatable)	Preventive and protective measures have failed to stop the process or relief the pressure and loss of containment has occurred.	Utilize mitigatory measures to contain the release and activate the emergency response measures to minimize the loss.

Table 5.2. Loss scenarios and applicable loss classes

Loss Scenario	Process Characteristic Region	Process State	Applicable Loss Classes	Applicable Loss Elements	Description of Loss Elements
Scenario 0	Central Region	Type 0	Class 0 (L_0)	l_0^O	Allowable operational loss
Scenario 1	Warning Region	Abnormal Type 1	Class I (L_I)	l_I^O	Unallowable operational loss
Scenario 2	Action Region	Abnormal Type 2	Class I (L_I)	l_I^O	Unallowable operational loss
			Class II (L_{II})	l_{II}^{BI}	Business interruption loss
Scenario 3	Mitigation Region	Abnormal Type 3	Class I (L_I)	l_I^O	Unallowable operational loss
			Class II (L_{II})	l_{II}^{BI}	Business interruption loss
				l_{II}^{R}	Reputational loss
			Class III (L_{III})	l_{III}^{A}	Asset loss
				l_{III}^{HH}	Human health loss
				l_{III}^{EC}	Environmental cleanup loss

5.2.2. Loss Scenarios

5.2.2.1. Loss Scenario 0: Normal Process

Under normal process conditions, it is assumed that the process characteristic is normally distributed with mean μ_0 and an initial variance σ_0^2. Deviations of the process characteristic, although within the specification limits, impose some degree of operational loss due to production of sub-quality products. However, the loss due to normal fluctuations of the process characteristic within the specification limits is considered as a constant characteristic of a normal process and assumed to be compatible with the acceptable tolerances. The loss Scenario 0 is quantified as:

$$L_{\text{Scenario 0}} = E^*\left\{l_0^O\left(\mu_0, \sigma_0\right)\right\} \times \int_0^t P(t)dt \tag{5.1}$$

where $P(t)$ is the production rate (e.g. barrel/hour) and $E^*\left\{l_0^O\left(\mu_0, \sigma_0\right)\right\}$ is the average observable loss per unit of produced product (e.g. dollar/barrel), referred to as the unit expected Class 0 loss in this work.

5.2.2.2. Loss Scenario 1: Abnormal Process Type 1

Under abnormal condition Type 1, it is assumed that at a random point in time the mean of the process characteristic changes in a time-dependent manner due to occurrence of an assignable cause. The change in variance is also assumed to occur at the same time as a change in the mean. According to Table 5.2, an abnormal process Type 1 experiences operational (Class I) loss due to process deviation beyond the specification limits. The cumulative expected operational loss for loss Scenario 1 is then determined as:

$$L_{\text{Scenario 1}} = \int_0^{\Delta\tau_C} P(t) E^* \left\{ l_I^O \left(\mu_t, \sigma_t \right) \right\} dt \tag{5.2}$$

where $P(t)$ is the production rate and $E^* \left\{ l_I^O \left(\mu_t, \sigma_t \right) \right\}$ is the unit expected Class I loss, which can be estimated as the expected value of MINLF. The $\Delta\tau_C$ is the length of time between the starting time of the abnormal process (τ_D) and the time when the process fault(s) is (are) detected and corrected (τ_C). The $\Delta\tau_C$ is assumed to be exponentially distributed with a mean of $\lambda_{\Delta\tau_C}^{-1}$.

In Equation (5.2), the μ_t and σ_t are the process characteristic mean and standard deviation at time t. Considering a deteriorating process, the general case includes an unstable process with a positive shift in mean with an age-dependent positive drift while process variance increases [9]. If $t < \tau_D$ (normal process) then $\sigma^2_t = \sigma_0^2$ and $\mu_t = \mu_0$. Alternatively, if $t > \tau_D$ (abnormal process) then $\sigma^2_t = \sigma_0^2 W_\sigma(t, \tau_D)$ and $\mu_t = \mu_0 + W_\mu(t, \tau)$, where $W_\sigma(t, \tau_D)$ and $W_\mu(t, \tau_D)$ represent respectively the variance and mean deteriorating function. In the simple case, a linear change in both mean and variance after maintaining a steady state for a while may be considered. In this case, the deteriorating functions are assumed as $W_\sigma(t, \tau_D) = (t - \tau_D + 1)$ and $W_\mu(t, \tau_D) = \delta_\mu + (t - \tau_D)\theta_\mu$ where δ_μ is the shift in mean, and θ_μ is the drift rate of mean. However, based on available information, any other deteriorating function that provides a better representation of the process deterioration can be used.

Finally, the distribution of the Class I loss is determined using the MC method by repeating the following two steps for J time:

Step 1. Simulate a realization of $\Delta\tau_C^j$ from a defined exponential distribution with a mean of $\lambda_{\Delta\tau_C}^{-1}$.

Step 2. Obtain an observation for loss Scenario 1 by replacing simulated $\Delta\tau_C^j$ in Equation (5.2).

From the loss distribution, the expected value of loss Scenario 1 can be obtained as the mean (or the median) of the simulated samples:

$$E\left\{L_{\text{Scenario 1}}\right\} = \frac{1}{J}\sum_{j=1}^{J} L_{\text{Scenario 1}}^{j} . \tag{5.3}$$

In the proposed approach, the greater amount of simulation runs (J) the better the precision of the estimation. The number of simulation runs of $J = 10^6$ is found as the value to ensure optimal balance between precision and computational time.

5.2.2.3. Loss Scenario 2: Abnormal Process Type 2

In Scenario 2, the operator fails to detect, diagnose, and correct the out-of-control process and the process fault(s) propagate(s). Over time the process characteristic deviates more from the target, the probabilities of unsafe conditions increase, and the process incurs more operational (Class I) loss. Once the process mean exceeds the high-high alarm (HHA) limit, it is considered that the successful activation of safety barriers shuts down the system at time $\tau_{SD.}$ The length of time during which the process experiences loss Scenario 2 is the summation of process downtime ($\Delta\tau_{dt}$) and business recovery ($\Delta\tau_{rp}$) periods. The overall loss associated with loss Scenario 2 is determined as:

$$L_{\text{Scenario 2}} = Aggregated\left\{L_{\text{I}}\,|_{0}^{\Delta\tau_{SD}}, L_{\text{II}}, L_{\varepsilon}\right\} \tag{5.4}$$

where $L_{\text{I}}\,|_{0}^{\Delta\tau_{SD}}$ captures the system operational loss over the period of time between deviation of process characteristic from its target value and the process shutdown ($\Delta\tau_{SD}$). L_{II} is the business loss and represents the lost production and associated business interruption losses during process downtime. A methodology to estimate the business interruption loss using the business interruption insurance (BII) approach is proposed in [1]. Zero reputational loss is considered for Scenario 2 as no loss of containment has occurred. The L_{ε} captures all other residual losses which are not covered by the defined loss classes. The loss aggregation methodologies are proposed in Section 5.4 based on the dependence structure among loss elements.

5.2.2.4. Loss Scenario 3: Abnormal Process Type 3

In Scenario 3, the failure of the control system and safety barriers causes unmanageable unsafe conditions. Failure to shut down the system or relieve the overpressure (if applicable) results in loss of containment (LOC). The cumulative overall loss associated with Scenario 3 is modeled as:

$$L_{\text{Scenario 3}} = Aggregated\left\{L_{\text{I}}\,|_{0}^{\Delta\tau_{LOC}}, L_{\text{II}}, L_{\text{III}}, L_{\varepsilon}\right\}. \tag{5.5}$$

The $\Delta\tau_{LOC}$ is the time span between the deviation of process characteristic and the loss of containment. L_I denotes operational loss class, L_{II} is the business loss class including the business interruption loss and reputational loss elements, and L_{III} represents event loss class

comprising asset loss, human health loss and environmental cleanup losses [1]. The application of copula functions in loss aggregation is discussed in the following sections.

5.3. Loss Aggregation Overview

5.3.1. Challenges in Loss Aggregation

Let L_1, L_2, …, L_D be the random variables that represent the elements of a given loss scenario i. Then, it can be shown that the expected value and the variance of the overall loss $L_{Scenario\ i} = L_1 + L_2 + \ldots + L_D$ are given by:

$$E\{L\} = \sum_{d=1}^{D} E\{L_d\}$$

and

$$Var\{L\} = \sum_{d=1}^{D} Var\{L_d\} + 2\sum_{d=1}^{D}\sum_{d'<d}^{D-1} Cov\{L_d, L_{d'}\}$$

where $E\{L_d\}$ and $Var\{L_d\}$ are, respectively, the mean value and the variance of each L_d, and $Cov\{L_d, L_{d'}\}$ is the covariance between L_d and $L_{d'}$. The covariance measures the dependence among random loss variables and has the value of zero for the case of independent losses. Thus, the $Var\{L\}$ is equal to the sum of the individual variances only for independent losses. As the covariance can be either positive or negative, for the case of dependent losses the $Var\{L\}$ can be either less than or higher than the sum of the individual variances. As it can be seen, dependence does not affect the mean overall loss. However, it does impact the variance significantly and will consequently affect the extreme values of the overall loss distribution (for example the 10^{th} and 90^{th} percentiles) [10].

Dependencies among loss elements within a single class or across the various loss classes are well recognized in the insurance and finance literature [11–13]. A comprehensive review of the consequence analysis literature related to process facilities shows that the potential dependencies among different losses have been virtually ignored in almost all of the existing methodologies. The superposition principle is used mostly to linearly add different loss elements to determine the overall loss [4,6,14–16]. Such an approach is applicable only if the losses for a given scenario are absolutely independent, which is not the case for most of the loss scenarios.

The dependence structure is recognized in this paper to estimate the overall loss. The next section describes a brief technical background of copulas to model the dependencies among random variables, which is used in Section 5.4 to develop a loss aggregation methodology.

5.3.2. Copula-Based Aggregation of Losses

5.3.2.1. Copula Functions

In the finance literature, there are several ongoing discussions about the application of copula models to account for possible dependencies among losses [11–13]. Copulas are used to describe the joint distribution of dependent random variables. With copula modeling, the marginal distributions from different families can be combined [17]. This is the main advantage of copulas compared with alternative methods, such as the use of multivariate distributions, to construct dependencies.

Let $\boldsymbol{L} = (L_1, ..., L_d)$ be a random vector with continuous marginal cumulative distribution functions (CDF) $F_{1,...,} F_d$. Based on Sklar's theorem, the CDF H of $\boldsymbol{L}$ can be represented as:

$$H(\boldsymbol{l}) = C\{F_1(l_1), \ldots, F_d(l_d)\}, \boldsymbol{l} \in \mathrm{R}^d,$$

in terms of a unique function C: $[0, 1]^d \rightarrow [0, 1]$, called a copula [18]. This representation suggests breaking the construction of a model for H into two parts: the estimation of the marginal CDFs $F_{1,...,} F_d$, and the estimation of the copula C. A comprehensive list of copula families and a review of the statistical issues involved in the model-building can be found Nelsen (2006).

5.3.2.2. Interactions and Copulas

To describe the dependencies among a set of related random variables, both linear and non-linear correlation coefficients could be used. The linear correlation coefficient expresses the linear dependence among normally distributed random variables. For non-linear cases, a rank correlation coefficient, such as the Kendall's τ or the Spearman's ρ, is more appropriate [19]. The possibility of using rank-correlation coefficients, which are insensitive to the marginal distributions, is a useful property of copula modeling. This advantage allows the random variables to interact and share information through the elements of the dependence matrix [20].

5.3.2.3. Selection of Copulas

Copula functions for a specific application may be based on actual observed data, or can be estimated using parametric methods [8], non-parametric methods and goodness-of-fit tests [21]. Alternatively, different copulas may be used to determine the sensitivity of simulation results to the input distribution.

Sectio 7.4 of this thesis provides a methodlogy to select the best fitting copula for a given application. A comprehensive review of the application of different copula-based methods used in economic risk modeling is provided elsewhere [7,11]. It is concluded that the Student-*t* copula (or simply *t*-copula) may be considered as the most appropriate copula to model the dependence structure for risk management as it is able to model both the center and tail dependencies of skewed distributions [7,11]. The *t*-copula is used in this work to construct the dependence structure among the loss elements and loss classes.

5.4. Loss Aggregation Methodology

Three different cases of loss aggregation are considered in this section:

- Case 1: Independent loss classes with independent loss elements
- Case 2: Independent loss classes with dependent loss elements
- Case 3: Dependent loss classes with dependent loss elements

In defining the above cases, it is assumed that the dependencies among loss classes are actually due to their dependent loss elements. Therefore, the case of dependent loss classes with independent loss elements is not considered. The loss aggregation approaches for the above cases are discussed in the following sections.

5.4.1. Case 1: Independent Loss Classes with Independent Loss Elements

The superposition principle for loss aggregation is used for Case 1 by simply adding all loss elements, assuming independent losses. The proposed approach is shown in Figure 5.2. For a given scenario, all applicable losses are added to estimate the overall loss. This step is repeated *J* times to determine the parameters of the overall loss distribution.

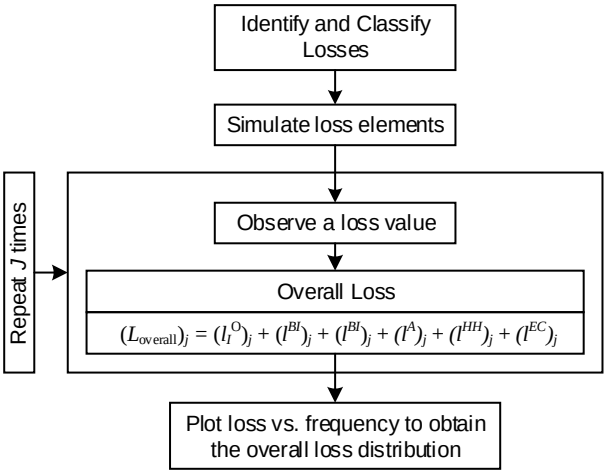


Figure 5.2. Case 1 loss aggregation using the superposition principle and Monte Carlo simulation

5.4.2. Case 2: Independent Loss Classes with Dependent Loss Elements

Copula functions are used to construct the dependence structure among elements of loss Classes II and III (note that Class I has only one element). Suppose there are CDFs $F_{L(h)}$, $h = 1, \ldots, H$ of loss values for each loss element. The distribution functions for the overall loss for each of the Classes II and III are obtained by repeating the following three steps for *J* times:

Step 1. Simulate a realization of a multivariate random vector $\bar{u} = (u_1, \ldots, u_H)$ with uniformly distributed marginal on [0, 1] from a defined copula *C*.

Step 2. Obtain a loss scenario for each loss element $l_j(h)$, $h = 1, \ldots, H$, by applying the inverse CDF to each uniformly distributed realization (simulated in previous step): $l_j(h) = F^{-1}_{L(h)}(u_h)$.

Step 3. Obtain a total loss scenario, L_j, summing losses $l_j(h)$ for each $h = 1, \ldots, H$:

$$L_j = \sum_{h=1}^{H} l_j(h).$$

The simulated empirical distribution functions for the overall loss for Classes II and III are obtained. Then, considering the independence of loss classes, the simulated random numbers of all loss classes are added to estimate the overall loss.

5.4.3. Case 3: Dependent Loss Classes with Dependent Loss Elements

In Case 3, the dependence structure is considered among all elements from all loss classes to provide a more realistic representation of a real-life situation. The loss aggregation approach for Case 3 is shown in Figure 5.3, which is the modification of the Case 2 loss aggregation by taking into account the correlation among all loss elements.

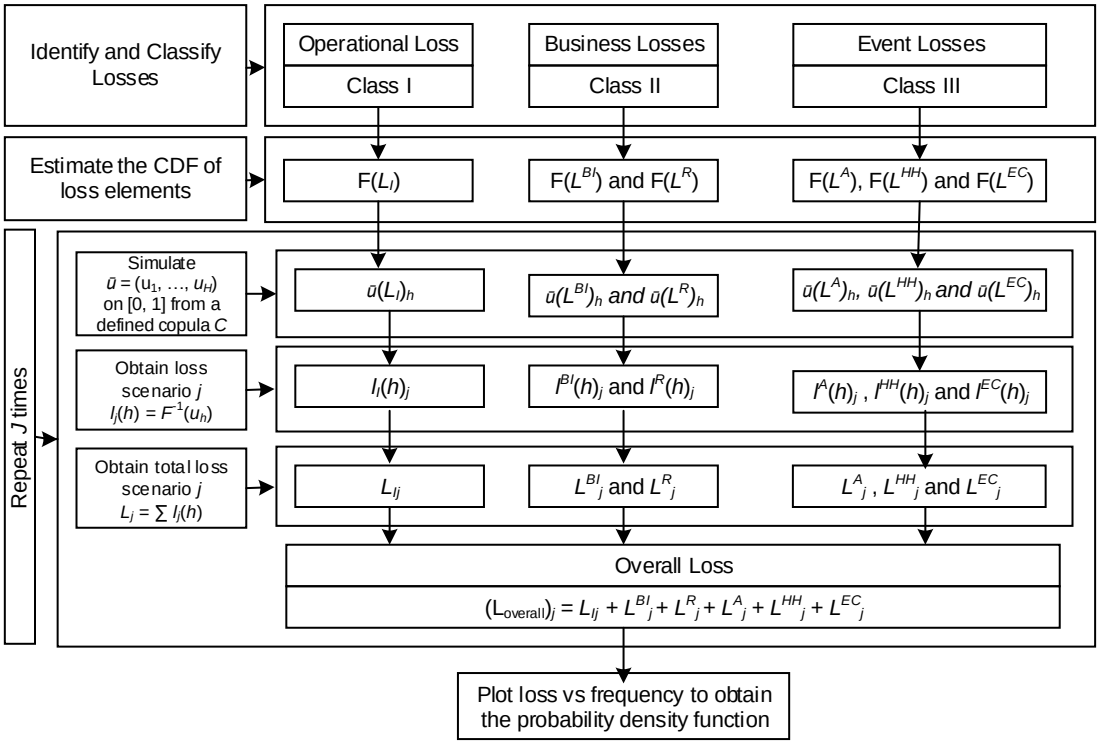


Figure 5.3. Case 3 loss aggregation: copula-based estimation of aggregated loss (dependent loss classes with dependent loss elements)

5.5. Case Study: Distillation Column

5.5.1. Process Description

The practical application of the proposed loss scenario analysis and loss aggregation methodologies is demonstrated using a hypothetical distillation column case study. Flooding is a common abnormal process condition that can cause loss of separation and negatively impact the safety and energy efficiency of the distillation process. Depending on the effectiveness of existing safety barriers, the system could experience different process end-states ranging from incipient flooding to runaway flooding [22].

To simplify the case study, assuming a fixed feed rate and pressure, the differential pressure (DP) across different sections of the column is used as the key process characteristic to measure the efficiency of the distillation process. A significant increase in DP across any section of a distillation column can indicate that the liquid level in the trays in that particular section is building, which is an indication of a potential flooding [23]. However, every process facility possesses a number of process characteristics that jointly impacts the process operational loss (e.g. temperature, DP, flow rate and product composition for a column). Development of multivariate loss functions for multiple dependent key process characteristics is a subject of an ongoing research by the authors.

Figure 5.4 shows the simplified event tree to identify different process states and loss scenarios due to DP deviation. For the sake of simplicity, control system, alarms, operator intervention, and emergency shutdown system (ESD) are considered as the only existing safety barriers. Based on the performance of these safety barriers, the column experiences different loss scenarios (Figure 5.4). The description of each loss scenario are presented in the following sections.

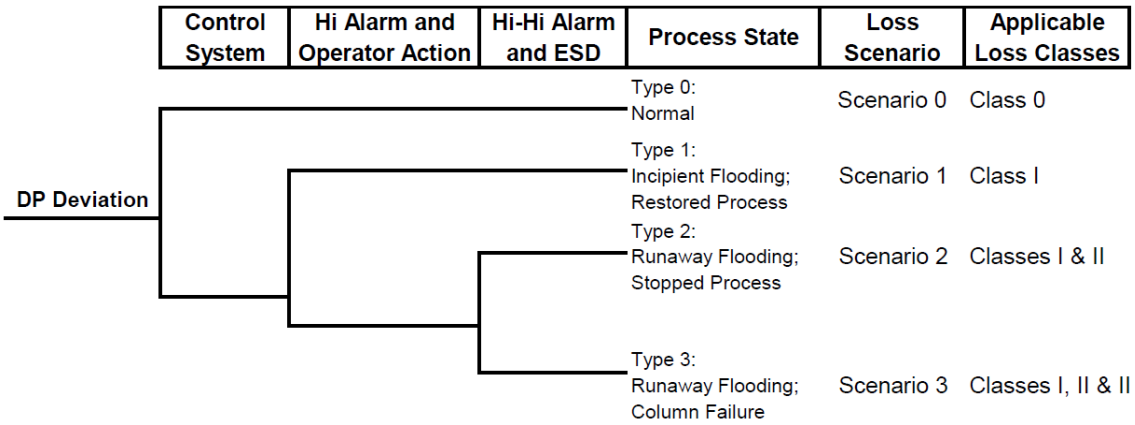


Figure 5.4. Event tree for DP deviations in the distillation column

5.5.2. Scenario 0: Normal Operation

During normal operation of the column, a normal distribution of the DP across the stripping section of the column with mean $\mu_0 = 2$ and inherent standard deviation $\sigma_0 = 0.25$ in inches of H_2O is assumed. The target DP is considered as $DP_T = 2$ inches of H_2O. The unit expected Class 0 operational loss is obtained as $E^*\left\{L_0^{O.P.}\right\}$ = \$0.89/bbl for column overpressure and $E^*\left\{L_0^{U.P.}\right\}$ = \$1.11/bbl for under-pressure scenarios. Then, using Equation (5.1) and the 20,000 bbl/day production rate, the accumulated Class 0 loss is obtained as $L_{\text{Scenario 0}}^{O.P.}$ = \$742/hour and $L_{\text{Scenario 0}}^{U.P.}$ = \$925/hour.

The calculated Class 0 loss is the loss due to inherent process variability resulting from control performance due to external disturbances and limitations on control action. It may be possible to reduce the variability and associated expected loss through improved control system performance, reduced measurement error, or advanced control functionality.

5.5.3. Scenario 1: Incipient Flooding—Restored Process

When the relative flow rates of the vapor and liquid are such that the drag force from the upward vapor flow is greater than or equal to the gravity force, the liquid stops flowing down the column. This condition is called incipient flooding [22]. The off-target operation of the column during flooding conditions causes the operational loss due to degradation of product quality. Equation (5.2) and the proposed MC method can be used to determine the Class I distribution.

5.5.4. Scenario 2: Runaway Flooding—Stopped Process

In Scenario 2, failure of the operator to correct the process fault triggers the runaway flooding. To avoid unsafe conditions and prevent damage to upstream equipment due to the column overflow, the high-high alarm and ESD system are implemented. This condition is referred to as abnormal process Type 2, where the system experiences operational (Class I) and business interruption (Class II) losses.

5.5.5. Scenario 3: Runaway Flooding—Loss of Containment

In Scenario 3, after failure of all existing safety barriers, a hypothetical incident in the distillation column is considered, where a runaway flooding resulted in column failure and material release. It is assumed that the flammable released hydrocarbons eventually found an ignition source, resulted in fire and explosion in the distillation unit. Deviation of the distillation column DP from its target value prior to the time τ_{LOC} caused operational loss primarily due to production of off-specification products and increased energy usage. During the distillation section shutdown, the organization's lost market share resulted in business interruption loss. Also, the attracted media attention and resulting negative public perception of the organization's safety performance and the lost market share imposed additional reputational loss. The loss of containment and subsequent fire and explosion caused event losses including human health loss, asset loss, and environmental cleanup loss.

Figure 5.5 shows the estimated distributions for different loss elements. The operational loss is estimated using the INLF. The business interruption insurance (BII) approach that

organizations use to propose business interruption claims to insurance companies is used to determine the business interruption loss. The Weibull distribution is used to represent the reputational loss, where the distribution parameters were estimated based on a scenario-based approach and management interviews. The component damage area is used to estimate the asset loss. The human health loss is calculated based on the estimated personnel injury consequence area along with the reported value of statistical life (VSL) in the literature. The environmental loss is considered as negligible assuming that the majority of the spill volume of fluid does not require cleanup due to rapid volatility. The distributions in Figure 5.5 are for illustration purposes; the calculation steps for each loss element are not within the scope of this work. Since the estimation of individual loss elements is not within the scope of this work, these calculations are skipped here to avoid repetition and to keep the paper concise. These estimated distributions are used in next section to demonstrate the application of the proposed loss aggregation methodology. The vertical dashed lines in Figure 5.5 show the 5^{th} percentile (P5), 50^{th} percentile (P50), and 95^{th} percentile (P95).

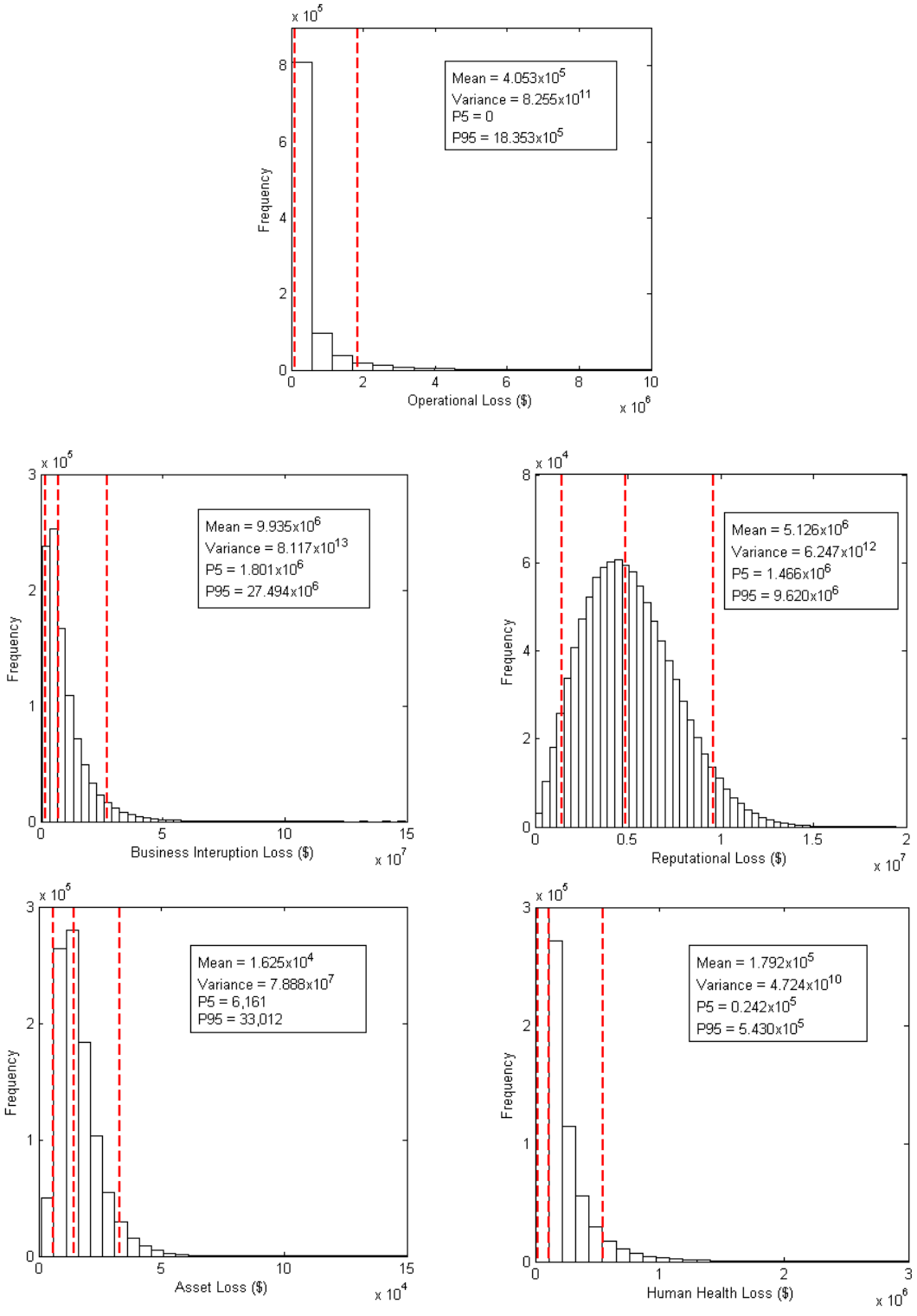


Figure 5.5. The distillation column loss distributions for Scenario 3. The vertical dashed lines show the P5, P50, and P95

5.5.6. Loss Aggregation

Having estimated the individual loss elements for each scenario, aggregation of losses is required for Scenarios 2 and 3 that have multiple loss elements in order to determine the overall loss. Scenario 3 is selected to demonstrate the application of the proposed loss aggregation methodologies. Similar aggregation steps are also applied to Scenario 2.

5.5.6.1. Case 1: Independent Loss Classes with Independent Loss Elements

The proposed numerical loss aggregation methodology (shown in Figure 5.2) is applied to loss Scenario 3 of the distillation column case study. For this purpose, the distributions of loss elements are integrated with the proposed MC method. The resulted histogram of the overall loss and the associated statistical information are shown in Figure 5.6 in Section 5.5.6.3. The three vertical lines in Figure 5.6 depict P10, P50, and P90.

5.5.6.2. Case 2: Independent Loss Classes with Dependent Loss Elements

The Case 2 loss aggregation methodology is used to simulate the dependency among the loss data within a loss class using the following three steps:

Step 1. Estimation of the marginal distributions for each loss element;

Step 2. Selecting the copula family and correlation parameters; and

Step 3. Determining the correlations among loss elements.

In Step 1, the marginal distributions can be estimated by fitting a parametric model separately to each loss element dataset. However, a parametric model may not be sufficiently flexible. Instead, an empirical model for the marginal distributions is used by

computing the empirical inverse CDF. Step 2 involves the selection of copula family and estimation of the correlation parameters. As discussed earlier, the *t*-copula is selected in this work to construct the dependence structure among loss elements.

In Step 3, the correlation coefficients are used to represent the degree of correlation in joint loss distributions using copulas. For this purpose, the Kendall's τ_k rank correlation coefficient is used, which is a function of the copula alone and is independent of the marginal distributions of the correlated loss elements. The closed-form expressions for τ_k rank correlation coefficient as a function of its correlating parameters (ρ) are provided in the literature [18]. Table 5.3 provides the specified ρ and associated τ_k rank correlation coefficients for the loss Scenario 3. Note that the correlation coefficients lie between 0 and 1, with higher values representing more interaction and lower values representing fewer interactions. The specifications in Table 5.3 are determined using expert knowledge. Based on available incident history and system loss performance data, these coefficients can be estimated using regression techniques to provide a better estimate of overall loss.

Table 5.3. Case 2 loss aggregation correlation coefficients

Loss Class	Loss Elements	Correlation Parameter (ρ)	Kendall's Coefficient (τ_k)
Class II	l_{II}^{BI} and l_{II}^{R}	$\rho\left(l_{II}^{BI}, l_{II}^{R}\right) = 0.75$	$\tau_k\left(l_{II}^{BI}, l_{II}^{R}\right) = 0.54$
Class III	l_{III}^{A} and l_{III}^{HH}	$\rho\left(l_{III}^{A}, l_{III}^{HH}\right) = 0.82$	$\tau_k\left(l_{III}^{A}, l_{III}^{HH}\right) = 0.61$

Using the three step technique discussed above, a *t*-copula, denoted by C_{ρ,ν_C}^{t}, is constructed between elements of Class II and III losses with the value of ρ specified in Table 5.3 and $\nu_C = 1$ (the degree of freedom). The developed Class II and Class III joint loss distributions

are then integrated with the MC method to estimate the Case 2 aggregated loss for the loss Scenario 3. The resulting histogram of the overall loss and the associated statistical information are shown in Figure 5.6. As can be seen from Figure 5.6, the mean and P50 of the overall loss are close to those of the Case 1; however, the P10 and variance are changed significantly. The variance of the overall loss in Case 2 is about 33% higher than in Case 1 due to the correlations considered among loss elements. In another words, ignorance of the positive correlations among loss elements underestimates the overall loss variance significantly. This observation is discussed further in Section 5.5.8 using a sensitivity analysis.

5.5.6.3. Case 3: Dependent Loss Classes with Dependent Loss Elements

In Case 3, the potential dependence among elements from different loss classes is also considered. Table 5.4 provides the specified correlation parameter for the Case 3 loss aggregation. The correlation parameters between elements of each loss class are assumed to be similar to those provided using expert knowledge in Table 5.3 for Case 2.

Using the same approach discussed for Case 2, a *t*-copula is constructed among the elements of loss classes. Figure 5.6 shows the histogram of the loss data with associated statistical information. As can be seen, the variance of overall loss data has a significant increase (11% higher than Case 2 and 48% higher than Case 1).

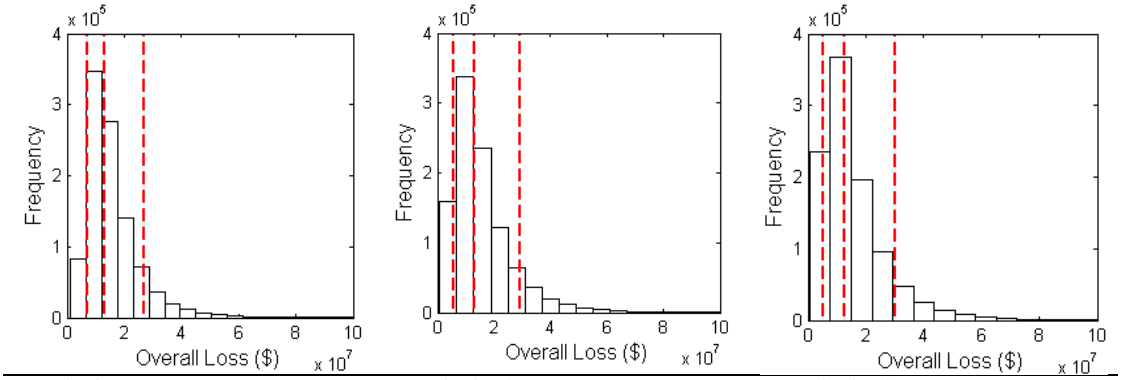
Table 5.4. Case 3 loss aggregation correlation coefficients

Correlation Parameter (ρ)

	l_I	l_{II}^{BI}	l_{II}^{R}	l_{III}^{A}	l_{III}^{HH}
l_I	1	0.55	0.25	0.45	0.35
l_{II}^{BI}	0.55	1	0.75	0.65	0.60
l_{II}^{R}	0.25	0.75	1	0.65	0.80
l_{III}^{A}	0.45	0.65	0.65	1	0.82
l_{III}^{HH}	0.35	0.60	0.80	0.82	1

Kendall's Coefficient (τ_K)

	l_I	l_{II}^{BI}	l_{II}^{R}	l_{III}^{A}	l_{III}^{HH}
l_I	1	0.37	0.16	0.30	0.23
l_{II}^{BI}	0.37	1	0.54	0.45	0.41
l_{II}^{R}	0.16	0.54	1	0.45	0.59
l_{III}^{A}	0.30	0.45	0.45	1	0.61
l_{III}^{HH}	0.23	0.41	0.59	0.61	1



	CASE 1	CASE 2	CASE 3
Mean	$15.735×10^6	$15.659×10^6	$15.732×10^6
Variance	8.812×10^13	11.766×10^13	13.085×10^13
P10	$7.116×10^6	$4.928×10^6	$5.377×10^6
P50	$13.341×10^6	$12.269×10^6	$12.719×10^6
P90	$27.144×10^6	$28.412×10^6	$29.596×10^6

Figure 5.6. Overall loss for distillation column loss Scenario 3. Left: Case 1—independent loss classes with independent loss elements; Middle: Case 2—independent loss classes with dependent loss elements; Right: Case 3—dependent loss classes with dependent loss elements

5.5.7. Discussion

Table 5.5 summarizes the results of the case study. The aggregated losses for Scenario 2 are estimated following the procedure described for Scenario 3. For the Scenario 2 Case 3 loss aggregation, the correlation coefficient between the operational loss and business interruption loss is estimated as $\rho\left(l_I^O, l_{II}^{BI}\right) = 0.45$ based on expert opinion.

Table 5.5 highlights the necessity of considering the dependence among loss elements. Although selection of copula function and estimation of correlation coefficients in Cases 2

and 3 require resources, these help to avoid underestimation of the loss variance. For practical applications, the Case 1 aggregation approach that considers independent losses can be used as a starting point. Then, by analysing the process business performance over time and the loss information determined from process incidents, the correlation coefficients among loss elements can be captured to implement Cases 2 and 3 aggregation approaches. To better understand the impact of dependency on the estimated overall loss distribution, a sensitivity analysis is conducted and results are presented in the subsequent section.

Table 5.5. Mean (μ) and variance (σ^2) of estimated loss values for the distillation column loss scenarios

	Estimated Loss Value		
Loss Type	**Scenario 1**	**Scenario 2**	**Scenario 3**
l_I^O	$\mu = \$3.05\times10^4$ $\sigma^2 = 4.70\times10^9$	$\mu = \$1.41\times10^5$ $\sigma^2 = 1.01\times10^{11}$	$\mu = \$4.07\times10^5$ $\sigma^2 = 8.32\times10^{11}$
l_{II}^{BI}	NA*	$\mu = \$2.83\times10^5$ $\sigma^2 = 1.32\times10^{10}$	$\mu = \$9.94\times10^6$ $\sigma^2 = 8.12\times10^{13}$
l_{II}^R	NA	NA	$\mu = \$5.13\times10^6$ $\sigma^2 = 6.25\times10^{12}$
l_{III}^A	NA	NA	$\mu = \$1.63\times10^4$ $\sigma^2 = 7.89\times10^7$
l_{III}^{HH}	NA	NA	$\mu = \$1.80\times10^5$ $\sigma^2 = 4.72\times10^{10}$
Overall Loss (Case 1†)	$\mu = \$3.05\times10^4$ $\sigma^2 = 4.70\times10^9$	$\mu = \$4.23\times10^5$ $\sigma^2 = 11.36\times10^9$	$\mu = \$15.74\times10^6$ $\sigma^2 = 8.81\times10^{13}$
Overall Loss (Case 2††)	NA	NA (Note 1)	$\mu = \$15.66\times10^6$ $\sigma^2 = 11.77\times10^{13}$
Overall Loss (Case 3‡)	NA	$\mu = \$4.23\times10^5$ $\sigma^2 = 15.08\times10^9$	$\mu = \$15.73\times10^6$ $\sigma^2 = 13.09\times10^{13}$

* NA: Not applicable
† Case 1: Independent loss classes with independent loss elements
†† Case 2: Independent loss classes with dependent loss elements
‡ Case 3: Dependent loss classes with dependent loss elements
Note 1: Not applicable because each class has only one element for Scenario 2

5.5.8. Sensitivity Analysis

A sensitivity analysis is performed to investigate the effect of the choice of correlation coefficient on the simulated overall loss for both positive and negative values of ρ. Table 5.6 shows the results of 11 experiments conducted with different ρ for *t*-copula with ν_C = 1 to determine the overall business (Class II) loss of Scenario 3. In each experiment, the MC method with $J = 10^6$ is used to generate the reputational and business interruption losses and the copula-based aggregation approach is applied to determine the overall Class II loss. Table 5.6 shows the P10, P50, and P90 percentiles along with the estimated variance. The variance ratios in Table 5.6 represent the ratio of the overall loss variance of the experiments with dependent losses to the variance of the overall loss in experiment 6 for independent losses ($\rho = 0$).

As can be seen from Table 5.6, as the value of ρ decreases from 0.95 to -0.95, the mean of the overall Class II loss remains almost constants; however, the variance, and therefore the variance ratios, decrease. The slight difference between the mean values is due to the sampling error, which can be reduced by increasing the number of simulation runs. Roughly speaking, the amount of dependence measures the degree to which large or small values of correlated losses associate with each other. In another words, dependence is different from overlapping structure among losses. While dependence affects the variance, the overlapping structure may affect the expected value of the overall loss. The study of overlapping structure could be a subject for further study, especially for the Class II loss where the lost market share may cause a potential overlap between reputational and business interruption loss elements.

Table 5.6 also shows the uncertainty factor (UF), defined as the ratio of P90 to P10 for different correlation coefficient values. The UF also decreases as the value of ρ decreases. Thus, it can be concluded that for positively correlated losses, which is the case for most process facilities operations, the uncertainty of the estimated overall loss is more sensitive to the uncertainty of model input parameters. Therefore, more research and resources will be required to choose less uncertain input parameters for the case of positively correlated losses, compared to independent and negatively correlated losses.

Table 5.6. The mean, variance ratio and percentiles of overall Class II loss for Scenario 3 for different ρ values

			Percentiles				Variance	
Experiment	**ρ**	**Mean**	**P10**	**P50**	**P90**	**Variance**	**Ratio**	**UF**
1	0.95	$\$15.06\times10^6$	$\$4.52\times10^6$	$\$12.11\times10^6$	$\$29.40\times10^6$	1.257×10^{14}	1.48	11.03
2	0.8	$\$15.13\times10^6$	$\$4.81\times10^6$	$\$12.26\times10^6$	$\$28.86\times10^6$	1.192×10^{14}	1.37	10.23
3	0.6	$\$15.10\times10^6$	$\$5.25\times10^6$	$\$12.33\times10^6$	$\$28.11\times10^6$	1.108×10^{14}	1.29	9.31
4	0.4	$\$15.13\times10^6$	$\$5.59\times10^6$	$\$12.44\times10^6$	$\$27.69\times10^6$	1.037×10^{14}	1.22	8.50
5	0.2	$\$15.11\times10^6$	$\$6.11\times10^6$	$\$12.48\times10^6$	$\$27.01\times10^6$	9.633×10^{13}	1.10	7.64
6	0	$\$15.01\times10^6$	$\$6.65\times10^6$	$\$12.53\times10^6$	$\$26.21\times10^6$	8.934×10^{13}	1.00	6.78
7	-0.2	$\$15.13\times10^6$	$\$7.34\times10^6$	$\$12.60\times10^6$	$\$25.84\times10^6$	8.288×10^{13}	0.94	5.96
8	-0.4	$\$15.04\times10^6$	$\$7.98\times10^6$	$\$12.57\times10^6$	$\$25.08\times10^6$	7.643×10^{13}	0.87	5.12
9	-0.6	$\$15.02\times10^6$	$\$8.77\times10^6$	$\$12.52\times10^6$	$\$24.39\times10^6$	6.809×10^{13}	0.79	4.29
10	-0.8	$\$15.00\times10^6$	$\$9.62\times10^6$	$\$12.43\times10^6$	$\$23.68\times10^6$	6.094×10^{13}	0.70	3.52
11	-0.95	$\$15.05\times10^6$	$\$10.39\times10^6$	$\$12.36\times10^6$	$\$23.27\times10^6$	5.393×10^{13}	0.61	2.95

5.6. Conclusions

The chemical process systems deteriorate and eventually move to an out-of-control state due to the occurrence of assignable causes. In the present study, the time-dependent process deterioration with non-constant mean and variance of process variables is studied. Methods are provided to estimate the time period at which a process may experience different loss

scenarios. The estimated time periods are required to quantify the overall magnitude of time-dependent losses such as operational and business losses.

The aggregation of multiple loss elements is also required to quantify the overall loss for scenarios with multiple loss elements. This paper considers three different cases of loss aggregation, including independent losses, dependent loss elements assuming independent loss classes, and dependent loss elements and classes. A copula-based loss aggregation methodology is proposed to consider the dependence structure among losses. Both linear aggregation of independent losses and a copula-based approach are applied in a distillation column case study. The *t*-copula is used to model the dependence structure due to its flexibility in modeling skewed distributions.

The results demonstrate the benefits of considering the correlation among losses to avoid misestimation of the overall loss variance. For the case of the distillation column case study, it is shown that ignoring the potential positive correlation among losses causes under-estimation of the overall loss variance by %48 compared to the simplified case of independent losses. The sensitivity of the approach to the choice of correlation parameters is investigated. The results show that increasing the correlation among loss elements does not affect the mean of the overall loss, however, it does affect the variance as the variance also increases significantly.

The case study results also highlight the potential for modelling errors if a wrong combination of copula function and correlation coefficient is selected. Therefore, it is imperative for the user to select the dependence structure that is most reflective of the system under analysis using the system's business performance and loss information from

process incident history. While the theoretical properties of copula functions are now fairly well understood, inference for copula models is, to an extent, still under development. The literature on the subject is yet to be collated in a future study, with a focus on choosing the best copula and estimating the copula parameters for application in process industries.

The paper demonstrates the benefits of copula modeling in estimating the overall loss. Although the implementation of the copula-based method adds to the complexity of the model, it enables a better description of the organization's loss exposure by taking into consideration the dependencies among losses. Integration of the Bayesian approach with the proposed loss models is another subject for future research to update the probability of different losses based on new information from the system. The ultimate goal is to develop a dynamic risk assessment methodology for process facilities by combining the proposed overall loss model in this paper with a dynamic probability analysis approach. The developed dynamic risk assessment tool would be more effective in generation of early warnings and accident predictions.

5.7. References

[1] Hashemi, S.J., Ahmed, S., Khan, F., 2014. Loss functions and their applications in process safety assessment. Process Saf. Prog. 33, 285–291. doi:10.1002/prs.11659.

[2] Rathnayaka S, Khan F, Amyotte P. SHIPP methodology: Predictive accident modeling approach. Part I: Methodology and model description. Process Saf Environ Prot 2011;89:151–64. doi:10.1016/j.psep.2011.01.002.

[3] Hashemi SJ, Ahmed S, Khan F. Risk-based operational performance analysis using loss functions. Chem Eng Sci 2014;116:99–108. doi:10.1016/j.ces.2014.04.042.

[4] API. Recommended Practice 581: Risk-Based Inspection Technology. 2nd ed. Washington: American Petroleum Institute; 2008.

[5] Pariyani A, Seider WD, Oktem UG, Soroush M. Dynamic Risk Analysis Using Alarm Databases to Improve Process Safety and Product Quality: Part I - Data Compaction. AIChE J 2012;58:812–25.

doi:10.1002/aic.12643.

[6] Arunraj NS, Maiti J. A methodology for overall consequence modeling in chemical industry. J Hazard Mater 2009;169:556–74. doi:10.1016/j.jhazmat.2009.03.133.

[7] Kole E, Koedijk K, Verbeek M. Selecting copulas for risk management. J Bank Financ 2007;31:2405–23. doi:10.1016/j.jbankfin.2006.09.010.

[8] Kolev N, Anjos U Dos, Mendes BVDM. Copulas: A Review and Recent Developments. Stoch Model 2006;22:617–60. doi:10.1080/15326340600878206.

[9] Tahera K, Ibrahim RN, Lochert PB. The effect of non-constant process variance in determining the optimal process means and production run of a deteriorating process. Prod Plan Control 2010;21:36–46. doi:10.1080/09537280903239433.

[10] Andrade DF, Barbetta PA, De Freitas Filho P, De Mello Zunino N, C. Jacinto C, Filho PJDF. Using Copulas in Risk Analysis. Proc. 2006 Winter Simul. Conf., Brazil: IEEE; 2006, p. 727–32. doi:10.1109/WSC.2006.323152.

[11] Soprano A, Crielaard B, Piancenza F, Ruspantini D. Measuring Operational and Reputational Risk – A Practitioner's Approach. Hoboken, NJ.: Wiley; 2010.

[12] Tang A, Valdez E a. Economic Capital and the Aggregation of Risks Using Copulas. SSRN Electron J 2009:1–29. doi:10.2139/ssrn.1347675.

[13] Temnov G, Warnung R. A Comparison of Loss Aggregation Methods for Operational Risk. J Oper Risk 2008;3:3–23.

[14] CCPS. Guidelines for Consequence Analysis of Chemical Releases. New York: Center for Chemical Process Safety of the American Institute of Chemical Engineers; 1999.

[15] Khan F, Amyotte PR. I2SI: A comprehensive quantitative tool for inherent safety and cost evaluation. J Loss Prev Process Ind 2005;18:310–26. doi:10.1016/j.jlp.2005.06.022.

[16] DNV. Recommended Practice DNV-RP-G101: Risk Based Inspection of Offshore Topsides Static Mechanical Equipment. October 20. Høvik, Norway: Det Norske Veritas (DNV); 2010.

[17] Genest C, Neslehová J. A Primer on Copulas for Count Data. Astin Bull 2007;37:475–515. doi:10.2143/AST.37.2.2024077.

[18] Nelsen RB. An Introduction to Copulas. 2nd ed. New-York: Springer–Verlag; 2006.

[19] MathWorks. Simulating Dependent Random Variables Using Copulas 2014. http://www.mathworks.com/help/stats/examples/simulating-dependent-random-variables-using-copulas.html (accessed May 15, 2014).

[20] Pariyani A, Seider WD, Oktem UG, Soroush M. Dynamic Risk Analysis Using Alarm Databases to Improve Process Safety and Product Quality: Part II - Bayesian Analysis. AIChE J 2012;58:826–41. doi:10.1002/aic.12642.

[21] Fermanian J-D. Goodness of fit tests for copulas. J Multivar Anal 2005;95:119–52.

[22] Rosemount. Distillation Column Flooding Diagnostics with Intelligent Differential Pressure Transmitter 2008. http://www2.emersonprocess.com/siteadmincenter/PM Rosemount Documents/3051S_ASP_Distillation_Column_Flooding.pdf (accessed April 10, 2014).

[23] Dzyacky GE. Process and apparatus for predicting and controlling flood and caryover conditions in a separation column. 5784538, 1998.

6. PROBABILISTIC MODELLING OF BUSINESS INTERRUPTION AND REPUTATIONAL LOSSES FOR PROCESS FACILITIES[5]

Preface

A version of this manuscript has been published in the Journal of Process Safety Progress. I am the primary author of this paper. Along with the co-authors, Faisal Khan and Salim Ahmed, I developed the conceptual model. I conducted the literature review and proposed alternate solutions to model business losses, including reputational loss. I prepared the first draft of the manuscript and subsequently revised the manuscript based on the co-authors' feedback and also the peer review process. The co-author Faisal Khan helped in developing and testing the concepts/models, reviewed and corrected the models and results, and contributed in preparing, reviewing and revising the manuscript. The co-author Salim Ahmed contributed through support in development, testing and improvement of the models. Salim Ahmed also assisted in reviewing and revising the manuscript.

Abstract

This paper presents probabilistic models to estimate business losses due to abnormal situations in process facilities. The main elements of business loss are identified as business interruption loss and reputational loss. The business interruption insurance approach is used to model business interruption loss. The sub-elements of business interruption loss

[5] Hashemi et al. Process Safety Progress 2015;34:373–82

are modeled based on expert knowledge using Program Evaluation Review Technique (PERT), which are then integrated using the Monte Carlo (MC) simulation approach. The reputational loss is considered as Weibull distributed and the parameters are estimated by applying a scenario-based approach. Copula functions are then used to develop the distribution of the aggregate loss, considering the correlation between business interruption and reputational losses. The application of the loss models is demonstrated using a distillation column case study. The models presented here provide a mechanism to monitor process facility's business performance, with associated uncertainties, and to make swift operational and safety decisions. This will help to improve process facilities safety performance and optimal allocation of resources where they are needed the most.

Keywords: Loss aggregation; loss modelling; copula; distillation column

6.1. Introduction

Continued occurrence of major losses in oil and energy sector along with adamant media attention as well as the strictly competitive and unstable oil market warn the oil and energy industry sector about the importance of managing business and reputational risks. Safe and cost effective performance of process facilitates is strongly correlated with the clients', regulators', and counterparties' perception of a company's trustworthiness [1–3]. The results of such perception, usually referred as company's reputation, not only affects the company in short term, but also affects growth sustainability in long term. Therefore, development of highly sophisticated process loss modelling methodologies has been an interesting area of research in recent years. Compared to accidental losses caused by the

release of material and energy, business and reputational loss modelling have received relatively less attention in the process industries literature. However, cost effective operation of process systems is equally important as the safety of operations, and both aspects require consideration for management decision making about continued operation and sustainable growth.

The business loss is defined in this work as the failure of an organization owning a process plant to generate enough revenue to cover all expenses associated with the process operation. The two most common causes for process systems to incur business loss are: (i) process shutdown due to process abnormal situation along with the failure of safety systems; and (ii) process downtime after an incident. A distinction has been made in this work between financial loss and business loss. Financial loss is a general term which refers to the expression of different losses (such as operational loss, asset loss, human health loss, environmental cleanup loss, as well as business loss) in monetary units. The objective of this work is to present methodologies for the modelling of process industries business losses.

Business loss is often difficult to quantify as it depends on several internal and external factors and loss is not simply attributed to the affected process unit as part of an overall service. The causes of business loss often have a severe impact on the organization in terms of business disruption and the service provided to clients, which eventually affects reputation [2]. Therefore, in the context of process facilities, business interruption loss and reputational loss are identified as the main elements of the business loss.

The existing loss modelling approaches for process industries are primarily based on fire, explosion, toxic release, and dispersion models [4–9]. The effects of reputational events are ignored in almost all existing models and studies. Moreover, in the existing methods to estimate the business interruption loss, only interrupted production and associated cost to restore the system are considered [4,5,10]. However, other relevant negative consequences and compensations such as loss of profit during the recovery period, insurance coverage, and fixed operating expenses should also be considered to avoid over/under-estimation of the financial consequences.

There could be a correlation between business interruption loss and reputational loss, which causes difficulty in separating them from each other. However, they are modeled separately in this work as different approaches are required to model each factor. The business interruption loss occurs mainly due to a gap in production, where extended downtime may also result in losing customers or market share, thus extending the loss of profit beyond production restart [11]. In contrast, the reputational loss occurs primarily due to the damaged organization's trustworthiness in the marketplace after a major incident with media coverage takes place [1,2].

Another contribution of this paper is to address the loss aggregation challenge, when different losses are correlated. The application of copula functions in process industries is reported recently as a promising tool for loss aggregation [12]. A copula-based approach is used in this work to estimate the aggregate business loss, considering the potential dependence structure among business interruption loss and reputational loss.

The organization of this paper is as follows. The modelling of business interruption loss and reputational loss are first discussed in Sections 6.6.2 and 6.6.3, respectively. Then a copula-based approach is provided in Section 6.6.4 to estimate the aggregate business loss. Finally, the practical application of the methodology is illustrated using a distillation process case study, followed by some concluding remarks.

6.2. Modelling of Business Interruption Loss

In most existing methodologies, the business interruption loss, denoted by L_{BI} in this work, is simply calculated based on production hours lost during process downtime multiplied by the production cost per hour [5,13]. Consideration of other expenses such as maintenance costs, material wastage costs, and material recycling costs are also proposed [4]. A rigorous method for estimating L_{BI} should take into account all major affecting factors to ensure proper evaluation [11].

To provide a better estimate of L_{BI}, this study applies the business interruption insurance (BII) approach that organizations use to propose claims to insurance companies [14,15]. Table 6.1 shows the elements of the proposed framework to calculate the expected L_{BI} after a process downtime.

Table 6.1. Elements of business interruption loss (L_{BI})

1. Profit loss due to lost production over process downtime in \$/day (l_{profit}^{dt})
2. Profit loss over recovery period in \$/day (l_{profit}^{rp})
3. All fixed operating expenses which will NOT be incurred during process downtime in \$/day (C_{fo}^{dt})
3.1. Regular maintenance costs
3.2. Building services
3.3. Utility bills
3.4. Rent (depends on lease agreement conditions during emergencies)
3.5. Expected payroll for staff who will not be employed during the indemnity period
3.6. Other
4. Expected costs to fix the damage and restore the business in \$ (C_d)
4.1. Start-up cost
4.2. Shutdown cost
4.3. Service costs (maintenance worker costs and contractor costs)
4.4. Material costs (spare parts acquisition, transportation, and spare parts inventory costs)
4.5. Material wastage and material recycling costs
4.6. Other
5. Business interruption insurance coverage (if applicable) in \$ (IC)

The first two elements in Table 6.1 measure the lost profit. Profit loss due to process downtime $\left(l_{profit}^{dt}\right)$ is to be determined as the expected gross revenues from sales of product over a period of time by projecting the past 12 to 24 months of the organization's sales forward, minus expected changes in inventory values, material use and transportation costs. Profit loss over the recovery period, denoted by l_{profit}^{rp}, occurs mainly due to lost market share and is determined by comparing the organization's business performance in the past 12 to 24 months before process downtime with the performance over the recovery period. If there is not enough past data available, for example for the case of a new business, the expert knowledge based on the data available from the process under study as well as data from similar processes can be used as a starting value to estimate profit loss. These loss

estimates should be revised as more information from the business performance becomes available over time.

The third element in Table 6.1 is fixed operating expenses (C_{fo}^{dt}) and is defined as those expenses which will not be incurred because of the process shutdown and include different sub-elements as shown in Table 6.1. The expected costs to fix the damage, the fourth element in Table 6.1 denoted by C_d, are the lump sum expenses associated with repairing the damaged facility. Finally, insurance coverage (*IC*) is defined as the percentage of the L_{BI} that will be recovered by the insurer.

Severity distribution is considered for the loss elements in Table 6.1 due to their stochastic nature. For this purpose, the severity is estimated through a distribution known as Program (or Project) Evaluation Review Technique (PERT). The PERT distribution is used in project and cost planning for modeling expert estimates of expected time, cost and other variables. The PERT distribution is a special case of the Beta distribution that uses three parameters: minimum, most likely (mode), and maximum values. Typically, sampling from the Beta distribution requires minimum and maximum values (scale) and two shape parameters, *v* and *w*. The PERT distribution uses the mode or most likely parameter to generate the shape parameters *v* and *w* [16].

In the proposed method, the expert is asked to estimate three values (minimum, most likely and maximum) for each element in Table 6.1. Then, a set of modified PERT distributions is plotted using Equation (6.1) and the expert is asked to select the shape that fits his/her opinion most accurately:

$$f(x) = \frac{(x - x_{\min})^{v-1}(x_{\max} - x)^{w-1}}{B(v,w)(x_{\max} - x_{\min})^{v+w-1}} \tag{6.1}$$

where x is any of the elements in Table 6.1 and $B(v, w)$ is a Beta function with parameters of the Beta distribution as:

$$v = 1 + \gamma_P \left(\frac{x_{\text{mode}} - x_{\min}}{x_{\max} - x_{\min}} \right) \tag{6.2}$$

and

$$w = 1 + \gamma_P \left(\frac{x_{\max} - x_{\text{mode}}}{x_{\max} - x_{\min}} \right). \tag{6.3}$$

The Equations (6.1) to (6.3) are used to determine PERT distributions for different elements in Table 6.1. In the standard PERT, γ_P equals 4. By increasing the value of γ_P, the distribution becomes progressively more peaked and concentrated around the mode (and therefore less uncertain). Conversely, by decreasing the γ_P the distribution becomes flatter and more uncertain [17]. The default γ_P = 4 can be used as the starting value to develop the PERT distributions.

By extending the estimated elements in Table 6.1 over the affected time, the overall business interruption loss is determined as:

$$l_{BI} = \Delta\tau_{dt}\left(l_{profit}^{dt} - C_{fo}^{dt}\right) + C_d + \Delta\tau_{rp} l_{profit}^{rp} - IC \tag{6.4}$$

where $\Delta\tau_{rp}$ and $\Delta\tau_{dt}$ are the recovery period and process downtime, respectively. Small letter l represents individual observations from a random loss variable L. Process downtime is defined as the following:

- For a stopped process with no loss of containment, process downtime is the time period between activation of the emergency shutdown device (ESD) due to a process fault and the diagnosis and correction of the fault.
- For a failed system that has experienced a release, process downtime is the time period between process shutdown and the time that the damaged equipment is repaired. An allowance should be made for the drawing up of new building plans, construction time, locating new premises, ordering, importing and installing new machinery.

The recovery period is the time span that the organization considers it would take from the production restart to restoring the business income to the same position it had before the loss occurred. An exponential distribution is considered for both $\Delta\tau_{dt}$ and $\Delta\tau_{rp}$. The exponential distribution is often used to model the time between events that happen with a constant occurrence rate at a random point in time. For the case of non-constant rates, distributions with a time-dependent hazard function, such as Weibull and lognormal, can be used [18]. The Dow Fire and Explosion Index is a typical method of estimating downtime after a fire or explosion [11], which can be used to estimate process downtime for applicable scenarios. Regarding the recovery period, when lacking information, $\Delta\tau_{rp} = 0.5\times\Delta\tau_{dt}$ can be used as a starting value. This value can be revised based on expert knowledge and failure history.

Having estimated the distribution of different elements in Equation (6.4), the Monte Carlo (MC) technique is used to model the distribution of overall L_{BI} by repeating the following steps for J realizations:

Step 1. Generate random variables representing each of the five elements in Table 6.1 from PERT distribution with the estimated parameters *v* and *w*.

Step 2. Generate random variables from exponential distribution to calculate $\Delta\tau_{dt}^{j}$ and $\Delta\tau_{rp}^{j}$.

Step 3. Determine the l_{BI}^{j} from Equation (6.4) and using loss elements simulated in Step 1 and $\Delta\tau_{dt}^{j}$ and $\Delta\tau_{rp}^{j}$ simulated in Step 2.

From the simulated loss distribution, the expected overall L_{BI} can be obtained as the mean (or the median) of the sample $\left[l_{BI}^{1},...,l_{BI}^{J}\right]$:

$$E\{L_{BI}\}=\frac{1}{J}\sum_{j=1}^{J}l_{BI}^{j}\ . \tag{6.5}$$

6.3. Modelling of Reputational Loss

The reputational loss is the missing element in almost all existing consequence analysis methods. There is no official definition of reputational loss and no loss measuring technique universally accepted by regulators or the industry. The reputational component is considered in this work as an intangible asset and an important part of a business involving a process facility. The effects of a reputational event may include, but are not limited to, the following [1–3]:

- Lost market share due to a change in the clients' preference to choose an alternate competitor.
- A fall in the company's share price.

- A negative impact on employees' confidence in the company's trustworthiness and difficulties in attracting highly-skilled human resources.
- An increased demand for greater disclosure and a need for quality control reassurance forced by regulators and rating agencies.
- For the case of major accidents, a negative impact of the entire sector's business.

The following section describes the methodology and the shortcomings of the share price volatility approach usually used in the literature to measure reputational loss. Then, Section 6.3.2 proposes an alternate scenario-based methodology to model reputational loss.

6.3.1. Share Price Volatility Approach

Soprano *et al.* [2] measured reputational loss of financial institutions as a function of the company's share price values. Way *et al.* [1] reviewed existing literature and also provided models to measure reputational loss as a function of share price volatility for process facilities. The assumption in share price volatility approach is that reputational events directly impact the company's market value [2]. One limitation of this approach is that it can only be applied for listed financial institutions. Moreover, the effect of insurance coverage is not considered in the share price volatility approach, which can result in an overestimation of the reputational loss. Other drawbacks of this approach include its dependence on availability of firm figures to represent the hard losses (such as products and assets, fines, and others) at specific points in time, and the difficulty to evolve the model into a predictive method [1]. Moreover, the results of a share price volatility

approach could be misleading as the share values might be affected by other risk factors and other forces at play, not only by the reputational component [2].

6.3.2. Scenario-Based Approach

An alternative scenario-based approach for modeling reputational loss (L_R) is applied in this paper to address the limitations of the share price volatility approach. The practical implementation of scenario-based approaches to measure the reputational loss has been reported in banking literature [2]. Figure 6.1 shows the framework for the scenario-based approach to measure reputational loss for process industries. As shown in Figure 6.1, firstly the critical processes which could suffer from reputational events should be identified. Then, for each critical process, the reputational scenarios are determined. A reputational scenario is defined as a hypothetical event that has not occurred but which could impact the organization. The next step includes carrying out management interviews using structured questionnaires. The minimum information that should be gathered for each identified scenario using interviews includes the average and variance of the loss amount and the potential insurance coverage. This information is then used to determine the expected reputational loss for each scenario.

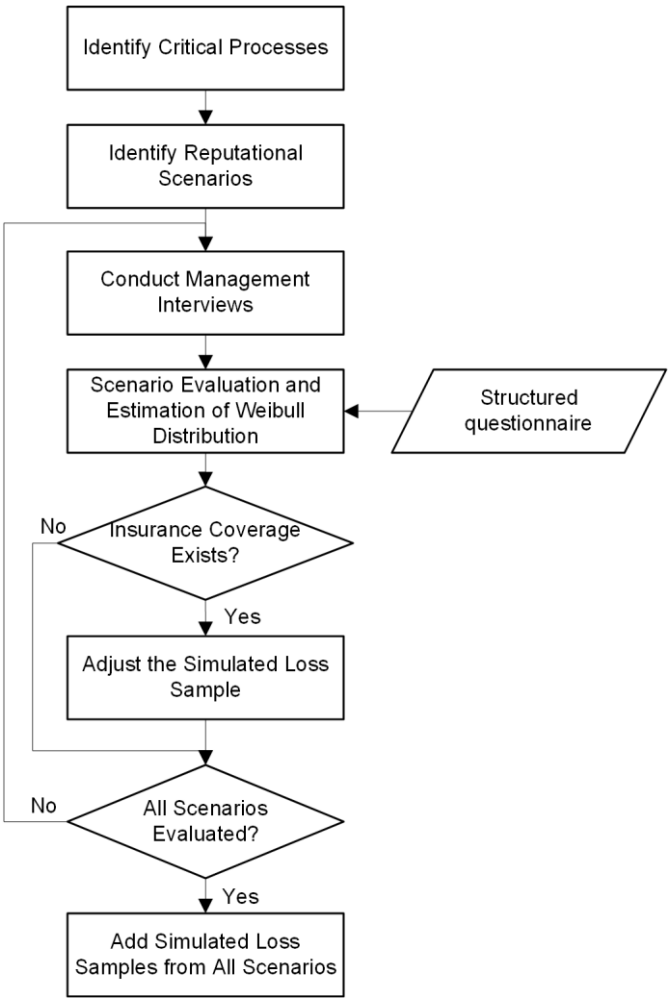


Figure 6.1. The proposed framework to estimate the reputational loss for process facilities

The severity of reputational loss is estimated through a Weibull distribution with the following probability density function [2]:

$$f\left(l_R\right)=\frac{\beta_w}{\theta_w}\left(l_R/\theta_w\right)^{\beta_w-1}\times\exp\left(-\left(l_R/\theta_w\right)^{\beta_w}\right) \tag{6.6}$$

where β_w and θ_w are the parameters of the Weibull distribution. The parameters of the distribution will be estimated using the empirical mean and the variance of loss data. The mean and the variance values, obtained from the questionnaire for management interviews,

are converted into the shape and scale parameters of the Weibull distribution using the following equations:

$$\overline{L_R} = \theta_w \cdot \Gamma\left(1+\frac{1}{\beta_w}\right), \tag{6.7}$$

$$\sigma^2_{L_R} = \theta^2_w\left[\Gamma\left(1+\frac{2}{\beta_w}\right)-\left(\Gamma\left(1+\frac{1}{\beta_w}\right)\right)^2\right] \tag{6.8}$$

where Γ is the gamma function. The loss distribution is simulated by applying the MC technique to draw *J* realizations from the severity distribution which results in a sample of simulated losses $\left[l_R^1,...,l_R^J\right]$. Protecting organizations against reputational losses is a relatively new service provided by some insurance companies [19]. Insurance coverage, if applicable, can be taken into account by applying an adjustment factor to the simulated loss samples. It is assumed that a fraction, α_{ins}, of the loss could be recovered ($0 \le \alpha_{ins} \le 1$); then the loss samples are adjusted as follows [2]:

$$L_{R(ins)} = \left[l_R^1 \cdot (1-\alpha_{ins}),...,l_R^J \cdot (1-\alpha_{ins})\right].$$

From the loss distribution, the expected reputational loss can be obtained as the mean (or the median) of the sample $\left[l_{R(ins)}^1,...,l_{R(ins)}^J\right]$:

$$E\left\{L_R\right\} = \frac{1}{J}\sum_{j=1}^{J} l_{R(ins)}^j \,. \tag{6.9}$$

If more than one scenario is identified, considering independence among scenarios, it is sufficient to add the simulated loss samples obtained from each scenario.

Due to the nature of the reputational loss and its strong correlation with other business loss elements, the application of a qualitative approach based on expert-knowledge is a more

viable option, compared to using quantitative methods like share price volatility. The main advantages of the interview-based approach to estimate the distribution parameters include, but are not limited to:

- Developing a questionnaire for interviews is flexible and easy to tailor to the process organization.
- Identifying areas of weakness and the need to focus on specific topics after a few interviews.
- The structure of the approach makes it easy for people to understand and participate in the interviews.

However, like any other expert knowledge-based analysis, the shortcoming of the proposed approach is the potential inconsistency of the assessment results. In order to tackle this challenge, each questionnaire can also be completed by managers, auditors, customers, and business specialists. Although this may add to the complexity of the assessment, this will make the assessment results robust and consistent. Proper training of the assessment team and all contributors involved in the reputational loss study is another key requirement to ensure successful implementation of the proposed method. Assigning an independent and trained reputational loss assessment team, along with a structured and auditable assessment procedure and management support, will ensure successful implementation of the proposed methodology.

6.4. Aggregate Business Loss

6.4.1. Copula-Based Loss Aggregation

The modeling of reputational loss (L_R) and business interruption losses (L_{BI}) is strictly related. The L_{BI} usually has a reputational impact and managerial decisions affecting L_R will impact business performance. Ignorance of the potential correlation between L_{BI} and L_R could result in misestimation of the aggregate loss [12]. The aggregation and combined management of losses are active areas of research in financial and banking literature [20–22], however, it has received less attention in the process industry's literature.

In an earlier work by the authors [12], the concept of copula functions is used to describe the joint distribution of dependent random loss variables and develop aggregate loss. With copula modeling, the marginal dependent losses with any distributions can be combined. Let L_{BI} and L_R be random variables representing business interruption loss and reputational loss with individual (marginal) loss distributions F and G and joint distribution L. Based on Sklar's theorem, the joint distribution L can be represented as:

$$L\left(l_{BI}, l_R\right) = C\left\{F\left(l_{BI}\right), G\left(l_R\right)\right\}, \tag{6.10}$$

in terms of a unique function C, called a copula [23], for all real values of l_{BI} an l_R. The main advantage provided to the risk practitioner by this representation is that the selection of an appropriate model for the dependence between losses, represented by C, can then proceed independently from the choice of the marginal loss distributions, F and G. For an introduction to the theory of copulas and a large selection of related models, the reader may refer, for example, to Nelsen [23], and an earlier work by the authors [12] in which process loss aggregation is considered.

6.4.2. Correlation and Copulas

When analyzing Gaussian data, means and variances–covariances could be modeled separately, and the dependency is uniquely characterized by Pearson's correlation coefficient that is intrinsically related with linear dependence and the normal distribution. For other types of data, in general, this is not valid anymore and a rank correlation coefficient, such as the Kendall's τ, is required [24]. Rank correlation τ takes values $\tau \in \langle -1,1 \rangle$, where -1 signals a perfect negative correlation, 1 displays a perfect positive correlation and 0 shows no correlation. The possibility of using rank-correlation coefficients, which are insensitive to the marginal distributions, is a useful property of copula modeling.

6.4.3. Copula Selection

It has been shown that the Student-*t* copula (or simply *t*-copula) may be considered as an appropriate copula to model the dependence structure for risk management and loss modelling as it is able to model both the center and tail dependencies of skewed distributions [2,22]. The *t*-copula is used in this work to construct the dependence structure among losses. Developing a framework to correctly and efficiently estimate the copula parameters and to discriminate between competing copula models for a specific application is a subject for an ongoing research by the authors.

6.4.4. Loss Aggregation Methodology

Consider $u_1 = F(l_{BI})$ and $u_2 = G(l_R)$, where F and G are the cumulative distribution functions (CDFs), associated to random loss variables L_{BI} and L_R. A t-copula is constructed between L_{BI} and L_R starting from a bivariate t distribution and degree of freedom ν_t, denoted by T_{τ,ν_t}, and transformation using the corresponding CDFs. Then, the developed bivariate t distribution is parameterized with the Kendall's rank correlation τ. Therefore, the joint business loss distribution determined from marginal distributions of L_{BI} and L_R using t-copula is:

$$L\left(l_{BI}, l_R\right) = C^t_{\tau,\nu_t}\left(F\left(l_{BI}\right), G\left(l_R\right)\right) = T_{\tau,\nu_t}\left(T^{-1}_{\nu_t}\left(u_1\right), T^{-1}_{\nu_t}\left(u_2\right)\right), \tag{6.11}$$

where C^t_{τ,ν_t} denotes a t-copula with Kendall's rank correlation τ and degree of freedom ν_t. Using the above procedure to construct dependence structure among losses, the aggregate loss distribution is then obtained by using the MC method and repeating the following three steps for J times:

Step 1. Simulate an observation from a bivariate random vector $\left(u_1^j, u_2^j\right)$ with uniformly distributed marginal on [0, 1] from a t-copula C^t_{τ,ν_t}.

Step 2. Apply the CDFs F^{-1} and G^{-1} to $\left(u_1^j, u_2^j\right)$ simulated in the previous step to determine a loss scenario as $l_{BI}\left(u_1^j\right) = F^{-1}\left(u_1^j\right)$ and $l_R\left(u_2^j\right) = G^{-1}\left(u_2^j\right)$.

Step 3. Obtain the total value of the loss scenario as: $L_j = l_{BI}\left(u_1^j\right) + l_R\left(u_2^j\right)$.

The empirical distribution of the business loss includes the simulated samples $\left[L_1, ..., L_J\right]$.

6.5. Case Study: Distillation Column

The proposed business loss modelling method is applied to a distillation column case study, as modified from Hashemi *et al.* [12]. In this case study, a fixed feed rate and pressure are assumed for the distillation column operation. Therefore, the differential pressure (DP) across the column can be used as the key process variable to monitor the operation of the distillation column. Figure 6.2 shows the safety barriers in place to prevent DP deviations and potential flooding conditions in the column. Failure of the safety barriers can cause the column DP to exceed a critical threshold that eventually triggers flooding conditions. The simplified event tree to identify different process end states due to DP deviation and associated loss scenarios are shown in Figure 6.2.

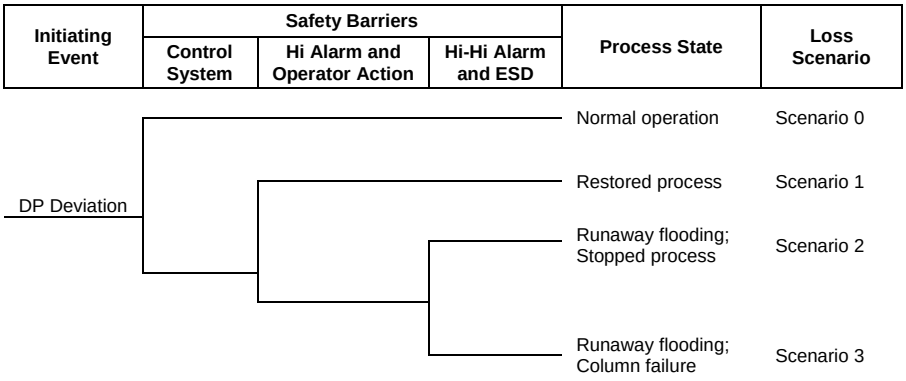


Figure 6.2. The simplified event tree for the distillation column case study

As shown in Figure 6.2, four different loss scenarios are identified for this case study based on the performance of the safety barriers. As no process downtime and reputational event have occurred in loss Scenario 0 (normal process) and Scenario 1 (restored process after a DP deviation) in Figure 6.2, zero business interruption loss is assumed for these two

situations. However, the system experiences business loss in Scenarios 2 and 3 due to the runaway flooding and the resulted process downtime. The description of loss Scenarios 2 and 3 and the calculation of the associated business interruption loss are provided in the following sections.

6.5.1. Scenario 2—Process Shutdown

In Scenario 2, upon failure of the control system and operator to correct the DP deviation, the high-high alarm and the emergency shutdown device (ESD) system are activated to shut down the process and avoid unsafe conditions. This process shutdown resulted in business loss due to lost production and associated expenses. However, as no major accident and reputational event have occurred, the business interruption loss (L_{BI}) due to stopped production is identified as the only element of the business loss.

The business interruption insurance (BII) questionnaire approach presented in Section 6.2 is used to determine the minimum, the most likely (mode) and the maximum values for different L_{BI} elements as identified in Table 6.1. These values are determined based on expert estimates by investigating the past 12 months of the organization's business performance and maintenance history and are presented in Table 6.2. For instance, the C_d in Table 6.2 (the total expected cost to correct the process fault and fix the damaged equipment after the process shutdown) is estimated using expert knowledge based on the recorded cost for similar previous maintenance activities. All other loss elements in Table 6.2 are estimated similarly. The parameters of Beta function (v_i and w_i) in Table 6.2 are determined from Equations (6.2) and (6.3).

Based on the process maintenance history, the process downtime after a shutdown is estimated using exponential distribution with $\lambda_{\Delta\tau_{dt}}^{-1}$ = 34 hours. From the organization's business performance, the mean of the business recovery period after a process shutdown due to activation of safety barriers is estimated as $\lambda_{\Delta\tau_{rp}}^{-1} = 0.5 \times \lambda_{\Delta\tau_{dt}}^{-1} = 17$ hours. Using the MC method proposed in Section 6.2 and Equations (6.4), the overall business interruption loss is simulated (see Figure 6.3 for the results).

Table 6.2. Business interruption loss elements for the distillation column case study. All monetary values are in thousand US dollars

		Scenario 2					Scenario 3				
Loss Element	**Description**	**Min.**	**Mode**	**Max.**	v_i	w_i	**Min.**	**Mode**	**Max.**	v_i	w_i
l_{profit}^{dt}	Profit loss due to production loss over process downtime (on a daily basis)	60	130	180	3.33	2.67	60	130	180	3.33	2.67
C_{fo}^{dt}	All fixed operating expenses which will not be incurred because of the process downtime (on a daily basis)	25	38	62	2.41	3.59	25	38	62	2.41	3.59
C_d	Total expected costs to fix the damage and restore the process	150	215	250	3.60	2.40	650	905	1,215	3.74	2.26
l_{profit}^{rp}	Profit loss over recovery period (on a daily basis)	5	8	11	3.00	3.00	14	25	39	2.76	3.24
IC	Business interruption insurance coverage	0	0	0	-	-	0	0	0	-	-

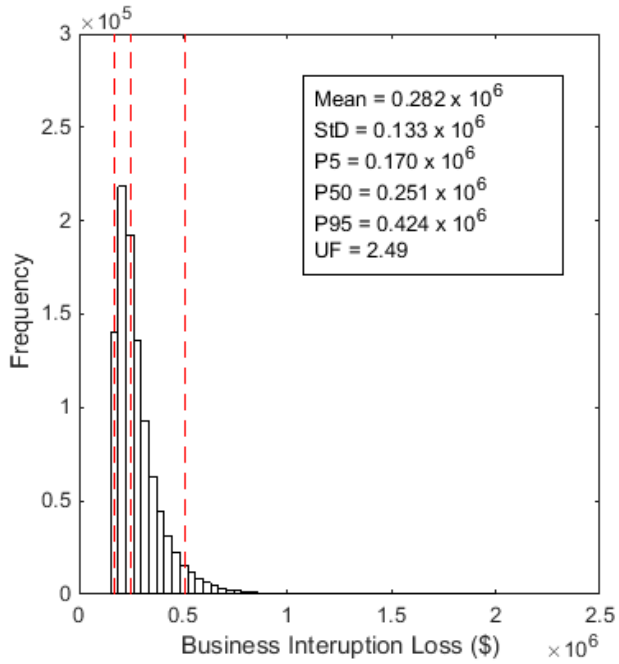


Figure 6.3. The distillation column business interruption loss for Scenario 2. The vertical dashed lines show the 5th percentile (P5), P50, and P95

6.5.2. Scenario 3—Loss of Containment

In Scenario 3, a hypothetical incident in the distillation column is considered, where runaway flooding resulted in the column failure and material release. It is assumed that the steam flow controller is failed and the reboiler heat duty started increasing at a constant rate. Increasing the reboiler heat duty caused more vapor to be boiled up, which increased the vapor flow up the column. Eventually, the column DP exceeded a critical threshold and triggered flooding. The operator failed to stop the runaway flooding and the column overflowed and sent liquid into the distillate process lines. When the hydrocarbon liquid overflowed into the outlet line of the column, the line ruptured due to mechanical shock. The hydrocarbon liquid and vapor mixture released from the outlet line became an

explosive mixture that drifted within the process area prior to being ignited by a heater. The incident included the fire in the crude distillation section and damage to several pieces of equipment and connecting pipelines with subsequent process shutdown.

During the distillation section shutdown, the production loss and consequent organization's lost market share resulted in business interruption loss. The attracted media attention and its negative impact on the public perception of the organization's safety performance imposed additional reputational loss. The estimation of the Scenario 3 L_{BI} and L_R are provided in the following sections. The loss of containment and subsequent fire and explosion caused accidental losses including human health loss, asset loss, and environmental cleanup loss. Estimation of these accidental losses is not within the scope of this paper.

6.5.2.1. Scenario 3—Business Interruption Loss

Table 6.2 shows the minimum, the most likely (mode), and the maximum values for different L_{BI} elements after the incident in the distillation column, determined based on expert estimates. As can be seen from Table 6.2, the expected cost to fix the damage and restore the process has a significantly higher value for Scenario 3 compared to Scenario 2 because of the extensive fire and explosion damage to the surrounding equipment in Scenario 3. The calculated Beta function parameters that are used to develop the PERT distributions are shown in Table 6.2. From the organization's business performance, the mean of the process downtime and business recovery period are estimated as $\lambda_{\Delta\tau_{dt}}^{-1} = 60$

and $\lambda_{\Delta\tau_{rp}}^{-1} = 120$ days, respectively. Figure 6.4 illustrates the results of the MC simulation for overall business interruption loss based on Equation (6.4) and with $J = 10^6$.

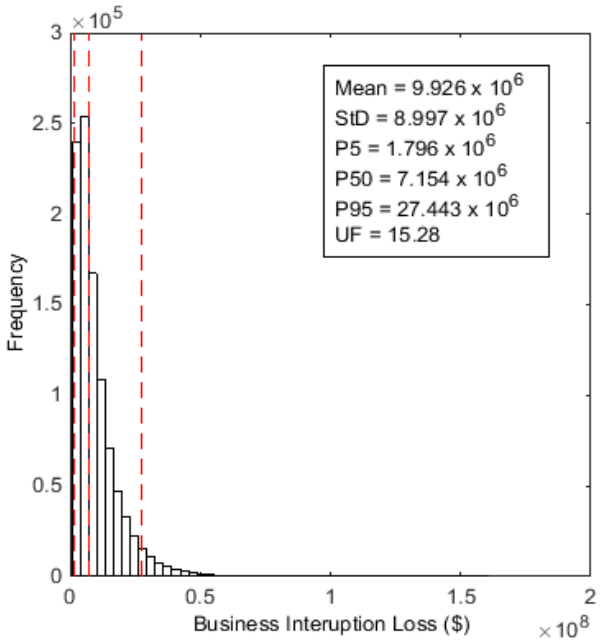


Figure 6.4. The distillation column business interruption loss for Scenario 3

6.5.2.2. Scenario 3—Reputational Loss

The severity of reputational loss (L_R) is estimated through a Weibull distribution using Equation (6.6). The empirical mean and variance of L_R are obtained from a questionnaire using hypothetical management interviews as $\overline{L_R} = 2.00\times10^6$ and $\sigma_{L_R}^2 = 0.16\times10^{12}$. Using Equations (6.7) and (6.8), these values are converted into the shape and scale parameters of the Weibull distribution as $\beta w = 5.79$ and $\theta w = \$2.16\times10^6$. After estimating the Weibull distribution parameters, the distribution of L_R is determined using the MC method with J

= 10^6 realizations. Zero reputational insurance coverage is assumed for this case study. Figure 6.5 shows the result of the MC simulation of L_R distribution.

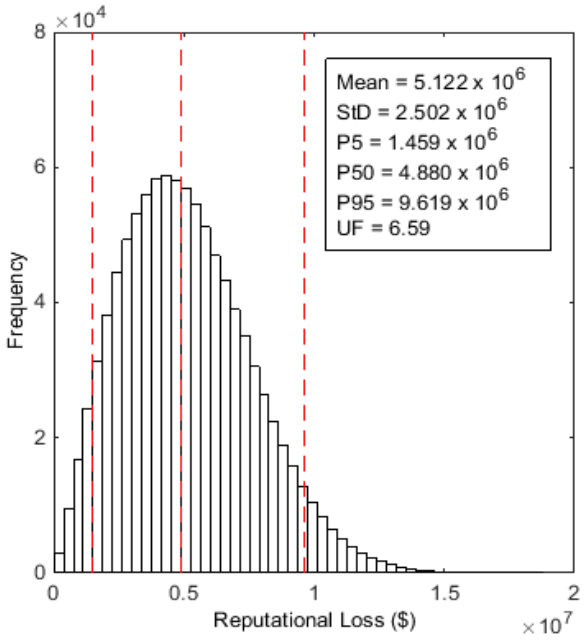


Figure 6.5. Reputational loss for distillation column case study Scenario 3

6.5.3. Uncertainty Analysis Methodology

This paper addresses the uncertainty in the proposed loss models through development of probability distributions for the input parameters and model outputs. For example, the application of the proposed PERT distributions aids in estimation of the distribution for subjective L_{BI} sub-elements based on expert estimates.

Having estimated different loss distributions, various percentiles of the distribution on the simulated loss can be extracted. The *n*th percentile (P*n*) is the level of loss that is exceeded with a probability of $(100-n)$ percent. The uncertainty is then characterized by the P5,

P50, and P95 of the uncertainty distribution on each of the loss distributions considered. The extent of the uncertainty is defined as the uncertainty factor (UF) and is presented using the ratio of the P95 to P5 percentiles [25]. The UF has a value equal or greater than 1, with smaller values of UF representing less uncertain estimated loss.

6.5.4. Discussion of Results

Table 6.3 summarizes the results of the case study, including the P5, P50, P95, and UF for loss elements of each scenario. The region between P5 and P95 of each loss is the area with 90% confidence of containing the true loss value. The UF size reflects the extent by which the loss model outputs may be affected by the input parameters values. The higher UF reflects the larger uncertainty associated with estimating the input elements of each loss model. This method of presenting the uncertainty shows the extent to which the uncertainty could be reduced by allocating resources required to better estimate the parameter values and to reduce the parameter uncertainties.

Table 6.3. The value of estimated loss elements for the distillation column (all monetary values are in million US dollars; UF is unitless)

	Scenario 2						Scenario 3					
Loss Element	**StD***	**Mean**	**P5**	**P50**	**P95**	**UF**	**StD**	**Mean**	**P5**	**P50**	**P95**	**UF**
l_{BI}	0.133	0.282	0.170	0.251	0.424	2.49	8.997	9.926	1.796	7.154	27.443	15.28
l_R	0	0	0	0	0	NA	2.502	5.122	1.459	4.880	9.619	6.59

* Note: StD denotes standard deviation.

As can be seen from Table 6.3, Scenario 2 with a process shutdown of a few days' length has considerably less business interruption loss amount in comparison to Scenario 3 which has several months of process downtime. A short process shutdown to correct the process fault and restore the process may not even be reflected in local media. These types of incidents have minimal reputational impact as a reasonable number of unplanned process shutdowns might be expected for a given process plant as long as they do not have any severe impact on people, properties and the environment. However, the loss of containment and subsequent fire and explosion in Scenario 3 with potentially severe consequences are usually reflected in national (or international) media with a negative impact on the reputation of the organization. Incidents similar to Scenario 3 involve several loss classes with different levels of severity, including business as well as accidental losses due to potential hazard scenarios. For such incidents, the loss models presented in this work can be used to model business loss elements. Accidental loss elements can be estimated using the well-established fire, explosion and dispersion models in the literature [4–9]. Then, the presented copula-based loss aggregation methodology in this work can be applied to estimate the overall loss of a given scenario.

6.5.5. Aggregate Business Loss

The presented copula-based loss aggregation methodology is applied to the loss Scenario 3 to estimate the aggregate business loss distribution by constructing the dependence structure between L_{BI} and L_R using the copula function. The marginal inverse CDFs required for Step 2 of the MC methodology in Section 6.4.4 can be estimated by fitting a

parametric model separately to each loss element dataset. However, a parametric model may not be sufficiently flexible. Instead, the empirical inverse CDFs for L_{BI} and L_R are computed from their empirical marginal distributions. Figure 6.6 shows the inverse CDF plots for L_{BI} and L_R in Scenario 3, determined from simulated loss datasets in Figures 6.4 and 6.5, respectively.

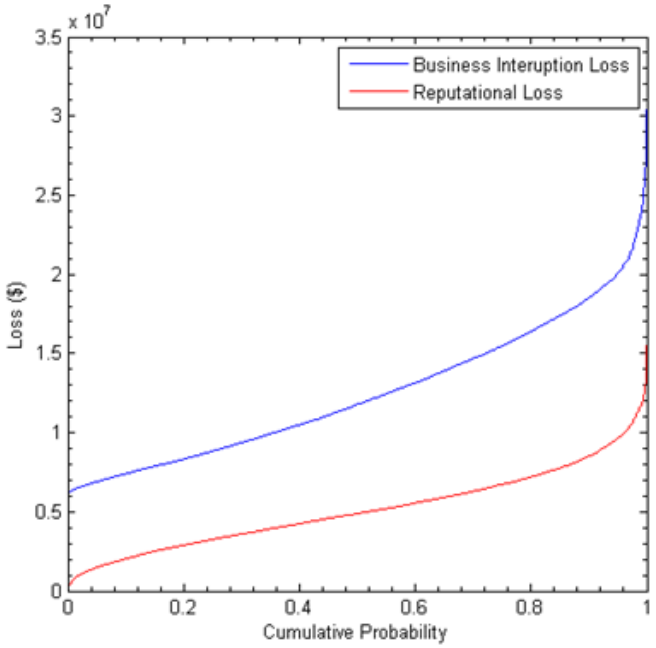


Figure 6.6. Inverse CDFs for loss Scenario 3

The next step involves a selection of the copula family and estimation of the rank correlation value. As discussed earlier, the *t*-copula is used in this work to construct the dependence structure among losses. The degree of freedom of $\nu_t = 1$ and the Kendall's τ value of 0.4 are determined using expert knowledge as the starting values. Based on available incident history and system loss performance data, these values can be revised

using the maximum likelihood evaluation method to provide a better estimate of joint loss distribution.

Figure 6.7 shows the joint loss distribution determined using the presented copula-based methodology. Figure 6.7 has histograms alongside a scatter plot to show both the marginal loss distributions and the dependence. It may be seen that the marginal histograms in Figure 6.7 closely match those of the original loss data in Figures 6.4 and 6.5 for L_{BI} and L_R, respectively.

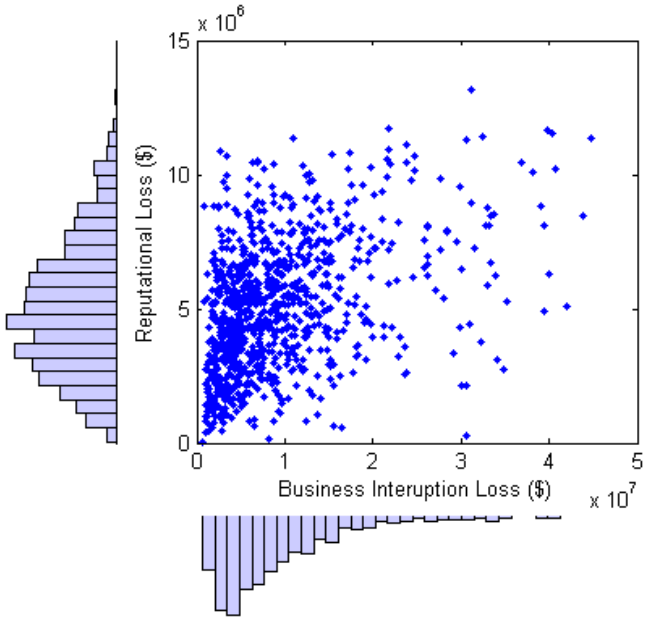


Figure 6.7. 1000 simulated business interruption loss and reputational loss data, using *t*-copula for loss Scenario 3

Table 6.4 shows the mean, standard deviation (*StD*) and different percentiles of the aggregate business loss distribution, determined from the proposed three-step MC simulation procedure in Section 6.4.4 for $J = 10^6$. To highlight the importance of

considering a dependence structure, Table 6.4 also shows the aggregate loss by assuming independence between L_{BI} and L_R and simply adding their values.

Table 6.4. The aggregate business loss for the distillation column loss Scenario 3. All monetary values are in million US dollars

Dependence Assumption	StD*	Mean	P5	P50	P95	UF
Simulated Using a *t*-copula $C^{t}_{\gamma=1,\tau=0.4}$	10.812	15.059	3.600	12.271	36.043	10.01
Independent Losses	9.348	15.060	5.352	12.663	33.005	6.17

As can be seen from Table 6.4, dependence structure between L_{BI} and L_R has almost no impact on the mean and P50 of the aggregate loss, but it does have a significant impact on the standard deviation, distribution extremes (e.g. P5 and P95 values), and the uncertainty factor. At first glance, it can be seen that putting effort to collect more information to construct the dependence structure between losses has increased the uncertainty of the aggregate loss. Although considering the positive dependence structure has increased the uncertainty, it provides a more realistic representation of the aggregate loss. In other words, ignoring the potential positive correlation among business loss elements, like most of the existing loss modelling methodologies, has the risk of under-estimating the aggregate loss uncertainty, leading to misallocation of resources.

6.6. Conclusions

In this study, models are proposed to assess the business loss elements, including business interruption loss and reputational loss. In addition to the cost of lost production and asset

repairs, the effects of lost market share during the recovery period, insurance coverage, and operating expenses are considered in the presented business interruption loss modelling. A scenario-based approach is proposed to model reputational loss that uses management interviews to identify reputational scenarios and the parameters of the loss distribution. The utilization of the MC method in the proposed loss models allows consideration of variability in input parameters and representation of loss results in the form of probabilistic distributions. The proposed approaches are flexible and can be adapted with respect to the level of details required. Depending upon the availability of the information, the user can simulate a model using either point-based estimates or probabilistic distributions for each variable. A mechanism is proposed to incorporate uncertainties in expert knowledge estimates using the PERT distributions and also to assess the uncertainty factor, which is determined as the ratio of 5th percentile to 95th percentile. This paper also demonstrates the flexibility and strength of copula-based approaches in modelling the dependence among losses while estimating the aggregate business loss.

The proposed models are applied to a distillation column. The results show the flexibility of the methodology in modeling business losses and associated uncertainties. The model presented for business interruption loss provides deeper understanding of the effects of contributing business loss elements, leading to an effective allocation of resources on actual loss drivers. Likewise, assessing and monitoring the reputational loss exposure of processes will improve relations with clients, investors, and regulators, and strengthen credit risk management, which will lead to lower losses and costs. The case study results also show that ignoring the potential positive correlation between estimated business interruption and

reputational losses causes under-estimation of the aggregate loss uncertainty. Therefore, care must be taken and resources should be allocated for proper estimation of the dependence structure among losses to avoid misestimating of the overall loss.

Integration of the Bayesian approach with the proposed loss models could be a subject for future research to update the loss distributions based on new loss information from the system. Future work by the authors will examine the application of the maximum likelihood evaluation method for copula parameter estimation and application of the information theory for best copula selection. The model is to be extended into a business risk assessment of process industries by considering both the frequency and severity distributions of losses.

6.7. References

[1] Way B, Khan F, Veitch B. Is Reputational Risk Quantifiable? Int. Conf. Mar. Saf. Environ. (IMSE 2013), Johor Bahru, Malaysia: 2013.

[2] Soprano A, Crielaard B, Piancenza F, Ruspantini D. Measuring Operational and Reputational Risk – A Practitioner's Approach. Hoboken, NJ.: Wiley; 2010.

[3] Condamin L, Louisot J, Naim P. Risk quantification: management, diagnosis and hedging. West Sussex, England: John Wiley & Sons, Inc.; 2006.

[4] Arunraj NS, Maiti J. A methodology for overall consequence modeling in chemical industry. J Hazard Mater 2009;169:556–74. doi:10.1016/j.jhazmat.2009.03.133.

[5] API. Recommended Practice 581: Risk-Based Inspection Technology. 2nd ed. Washington: American Petroleum Institute; 2008.

[6] CCPS. Guidelines for Consequence Analysis of Chemical Releases. New York: Center for Chemical Process Safety of the American Institute of Chemical Engineers; 1999.

[7] Pula R, Khan FI, Veitch B, Amyotte PR. A Grid Based Approach for Fire and Explosion Consequence Analysis. Process Saf Environ Prot 2006;84:79–91. doi:10.1205/psep.05063.

[8] Hashemi SJ, Ahmed S, Khan F. Risk-based operational performance analysis using loss functions. Chem Eng Sci 2014;116:99–108. doi:10.1016/j.ces.2014.04.042.

[9] Hashemi SJ, Ahmed S, Khan F. Loss functions and their applications in process safety assessment. Process Saf Prog 2014;33:285–91. doi:10.1002/prs.11659.

[10] Krishnasamy L, Khan F, Haddara M. Development of a risk-based maintenance (RBM) strategy for a power-generating plant. J Loss Prev Process Ind 2005;18:69–81. doi:10.1016/j.jlp.2005.01.002.

[11] API. Recommended Practice 580: Risk-Based Inspection. 2nd ed. Washington: American Petroleum Institute; 2009.

[12] Hashemi SJ, Ahmed S, Khan FI. Loss scenario analysis and loss aggregation for process facilities. Chem Eng Sci 2015;128:119–29. doi:10.1016/j.ces.2015.01.061.

[13] Khan F, Amyotte PR. I2SI: A comprehensive quantitative tool for inherent safety and cost evaluation. J Loss Prev Process Ind 2005;18:310–26. doi:10.1016/j.jlp.2005.06.022.

[14] Dominion Insurance. Business Interruption Proposal 2014. http://www.dominioninsurance.to/download/DominionTongaBusinessInterruptionProposal24.06.07.pdf (accessed March 23, 2014).

[15] Claims Canada. 10 Steps to Settling Business Interruption Claims 2008. http://www.claimscanada.ca/issues/article.aspx?aid=1000225358&er=NA (accessed April 2, 2014).

[16] RiskAMP. The beta-PERT Distribution 2014. http://www.riskamp.com/beta-pert#the-pert-distribution (accessed May 4, 2014).

[17] Vose D. Risk Analysis – A Quantitative Guide. 3rd ed. West Sussex, England: John Wiley & Sons, Inc.; 2008.

[18] Ebeling CE. An introduction to reliability and maintainability engineering. 2nd ed. Waveland Press, Inc.; 2010.

[19] ACE European Group. Reputation at Risk: ACE European Risk Briefing 2013 2013. http://www.acegroup.com/global-assets/documents/Europe-Corporate/Thought-Leadership/ace_reputation_at_risk_july_2013.pdf (accessed July 24, 2014).

[20] Bouy E, Roncalli T. Copulas for Finance A Reading Guide and Some Applications 2000.

[21] Kolev N, Paiva D. Copula-Based Regression Models 2007.

[22] Kole E, Koedijk K, Verbeek M. Selecting copulas for risk management. J Bank Financ 2007;31:2405–23. doi:10.1016/j.jbankfin.2006.09.010.

[23] Nelsen RB. An Introduction to Copulas. 2nd ed. New-York: Springer–Verlag; 2006.

[24] Accioly R, Chiyoshi FY. Modeling dependence with copulas: a useful tool for field development decision process. J Pet Sci Eng 2004;44:83–91. doi:10.1016/j.petrol.2004.02.007.

[25] Jones JA, Ehrhardt J, Goossens LHJ, Brown J, Cooke RM, Fischer F, et al. Probabilistic Accident Consequence Uncertainty Assessment Using COSYMA: Methodology and Processing Techniques. 2001.

7. OPERATIONAL LOSS MODELLING FOR PROCESS FACILITIES USING MULTIVARIATE LOSS FUNCTIONS[6]

Preface

A version of this manuscript has been published in the Journal of Chemical Engineering Research and Design. I am the primary author of this paper. Along with the co-authors, Faisal Khan and Salim Ahmed, I developed the conceptual model and subsequently translated this to the numerical model. I conducted the literature review and proposed a new technique to develop multivariate loss functions using copulas. I carried out most of the data collection and the comparison of loss functions. I prepared the first draft of the manuscript and subsequently revised the manuscript based on the co-authors' feedback and also the peer review process. The co-author Faisal Khan helped in developing and testing the concepts/models, reviewed and corrected the models and results, and contributed in preparing, reviewing and revising the manuscript. The co-author Salim Ahmed contributed through support in the development, testing and improvement of the models. Salim Ahmed also assisted in reviewing and revising the manuscript.

[6] Hashemi et al. Chemical Engineering Research and Design 2015;104:333–45.

Abstract

This paper presents a methodology to develop multivariate loss functions to measure the operational loss of process facilities. The proposed methodology uses loss functions to provide a model for operational loss due to deviation of key process characteristics from their target values. Having estimated the marginal loss functions for each monitored process variable, copula functions are then used to link the univariate margins and develop the multivariate loss function. The maximum likelihood evaluation method is used to estimate the copula parameters. Akaike's information criterion (AIC) is then applied to rank the copula models and choose the best fitting copula. A simulation study is provided to demonstrate the efficiency of the copula estimation procedure. The flexibility of the proposed approach in using any form of the symmetrical and asymmetrical loss functions and the practical application of the methodology are illustrated using a separation column case study.

Key words: Process risk assessment; copula function; multivariate model; process safety; distillation column; Akaike's information criterion (AIC)

7.1. Introduction

Different sources of variations in a process operation, such as feed specifications, wrong settings, control system malfunction and operator error can cause deviation of process variables from the specification limits. The subsequent unprofitable process operation incurs operational loss, which is defined in this work as the loss due to production of sub-quality products and increased energy usage resulting from a deviated process variable.

Process facilities possess different characteristics that jointly impact process operational loss. For example, the temperature and differential pressure across a distillation column can be used jointly to monitor the operational loss of the distillation system. Thus, integrated operational loss modeling of process industries requires understanding the joint distribution of all key process characteristics and their correlations.

The loss function approach is widely used to quantify quality loss in the manufacturing industry [1,2] by relating a key characteristic of a system (e.g. product composition) to its business performance. More recently, loss functions have been applied to model operational loss for process facilities [3]. Choosing and estimating a useful form for the marginal loss functions of each process characteristic is often a straightforward task [3,4], given that enough loss information from the system is available. For multivariate cases, traditionally, the pairwise dependence between loss functions has been described using traditional families of loss functions. The two most common models occurring in this context are the multivariate quadratic loss function (QLF) [5,6] and the multivariate inverted normal loss function (INLF) [7,8]. For instance, Spiring [7] proposed the following equation for bivariate cases with two parameters for which INLF can be used to represent operational loss:

$$L(\boldsymbol{Y}) = EML\left[1-\exp\left\{-\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{T})^{\mathrm{T}}\,\boldsymbol{\Gamma}^{-1}(\boldsymbol{Y}-\boldsymbol{T})\right\}\right] \tag{7.1}$$

where $\boldsymbol{Y}$ and $\boldsymbol{T}$ are 2 × 1 column vectors of key process characteristics under scrutiny and associated target values, respectively. *EML* is the maximum estimated loss and $\boldsymbol{\Gamma}$ is a 2 × 2 scaling matrix (shape parameter) relating deviation from target to loss for both

parameters. The main limitation of this approach is that the individual behaviour of the marginal loss functions must then be characterized by the same parametric family of loss functions. This restriction has limited their useful application in practical situations. Moreover, other than the QLF and INLF, loss functions usually do not have a convenient multivariate generalization.

According to a review of the existing literature in the area of multivariate loss functions conducted by Hashemi et al. [3], it can be concluded that the existing research challenge is to develop a flexible framework to assign appropriate marginal loss functions to key process characteristics irrespective of their dependence structure. Copula models, which provide this flexibility, have begun to make their way into process engineering literature [9–11]. Copulas are used to describe the joint distribution of dependent random variables with any marginal distribution. While the theoretical properties of copula functions are now fairly well understood, inference for copula models is, to an extent, still under development [12].

The contributions of this paper are twofold. First, a new methodology is provided to construct multivariate loss functions using copulas. Second, methodologies are provided to estimate copula parameters and choose the best copula for a specific application. The main objective of this paper is to present the successive steps required to use copulas for modelling the dependent losses and constructing multivariate distributions for specific purposes, including operational loss modelling.

Following the introduction, Section 7.2 proposes a methodology to develop multivariate loss functions using copula functions. Section 7.3 reviews the theory of copula functions

and Section 7.4 provides methods to estimate and select copula functions and conduct an uncertainty assessment. A separation column case study is then used in Section 7.5 to illustrate the practical implementation of copulas, followed by some concluding remarks.

7.2. Methodology: Copula-Based Multivariate Loss Functions

It has been shown in earlier studies that the application of the general class of inverted probability loss functions (IPLFs) is a flexible approach to model loss due to process deviations [1,3]. However, the application of IPLFs for systems with multiple key process variables is an existing research challenge due to the restriction in multivariate generalizations. Copula functions are used in this work to overcome this challenge.

Before developing a multivariate loss function, it would be helpful to review the common basis of developing IPLFs. According to [1], let $f\left(x_i\right)$ be a probability density function (PDF) possessing a unique maximum at x_i, where x_i represents a key process characteristic (KPC) and $i = 1, \ldots, I$ represents I KPCs (e.g. temperature, pressure, composition, etc.). Let $T_i = x_i$ be the value at which the PDF attains its unique maximum, where denotes the target value. Let $\pi\left(x_i, T_i\right) = f\left(x_i\right)$, $m_i = \sup_{x_i \in X_i} f\left(x\right) = f\left(T\right)$, and define loss inversion ratio (LIR) as:

$$f_{LIR}\left(x_i, T_i\right) = \pi\left(x_i, T_i\right) / m_i . \tag{7.2}$$

Then, any IPLF takes the form:

$$L\left(x_i, T_i\right) = EML_i\left[1 - \pi\left(x_i, T_i\right) / m_i\right] \tag{7.3}$$

where EML_i is the maximum estimated loss incurred when the target is not attained. It can be seen from the structure of Equation (7.3) that $\pi(x_i, T_i)$ is in the form of a PDF in terms of x_i and T_i, m_i is the maximum of $\pi(x_i, T_i)$; the LIR, $\pi(x_i, T_i)/m_i$, is unitless and has a minimum value of zero when x_i takes on values far from the T_i, and a maximum value of one when x_i is exactly on target, i.e., $0 \le \pi(x_i, T_i)/m_i \le 1$ [1].

Table 7.1 shows the important IPLFs determined from inversion of Normal, Gamma, and Beta distributions using the method described above. A comparative study of the flexibility of different IPLFs for application in the process industries is provided in [3].

Table 7.1. Listing of univariate inverted probability loss functions (IPLFs)

Type of Loss Functions	Reference	Formulation of Loss Function†
INLF	[7]	$L(x,T) = EML\{1-exp(-(x-T)^2/2\gamma^2)\}$ where $\gamma = \Delta_x/4$
Modified INLF	[13]	$L(x,T) = \frac{EML_\Delta}{1-exp\{-0.5(\Delta_x/\gamma)^2\}}\{1-exp(-(x-T)^2/2\gamma^2)\}$
IBLF	[14]	$L(x,T) = EML\{1-D[x(1-x)^{(1-T)/T}]^{(\alpha-1)}\}$ where $D = [T(1-T)^{1-T/T}]^{1-\alpha}$
IGLF	[1]	$L(x,T) = EML\left\{1-(e/T)x\exp(-x/T)^{\alpha-1}\right\}$

† EML_Δ is the estimated maximum loss at distance Δ_x, where Δ_x is the distance from the target to the point where the maximum loss *EML* first occurs; x represents the process measurement; T denotes the target value; γ and α are shape parameters.

The same basis as in Equation (7.3) is used in this work to develop multivariate loss functions. As shown in Figure 7.1, the proposed methodology includes the following steps:

Step 1a) The proposed methodology starts with the identification of key process characteristics (KPCs), x_i, $i = 1, \ldots, I$. A KPC is a feature that, if nonconforming, missing, or degraded, may cause unsafe conditions and/or a loss of product

quality. For example, operating temperature is the KPC for a polymerization reactor. Different approaches, such as check lists, preliminary hazard analysis (PHA), failure modes and effects analysis (FMEA), fault tree analysis (FTA), hazard and operability study (HAZOP), and master logic diagrams, are often used to identify KPCs [4]. In this study, it is assumed that the KPCs are known.

Step 1b) The next step is to assign a loss inversion ratio, $LIR_i = \pi\left(x_i, T_i\right)/m_i$, to each identified KPC. A least-squares based method to determine the parameters of each LIR is described in Hashemi et al., (2014a).

Step 2) The best copula function and associated parameter(s) should then be selected to represent the dependence structure among identified LIRs.

Step 3) The copula decomposition property (see Equation (7.3) in Section 7.3) is then used to develop the multivariate density (MVD) function from the product of copula PDF and marginal LIRs. Finally, the multivariate loss function is developed by inverting the multivariate density function.

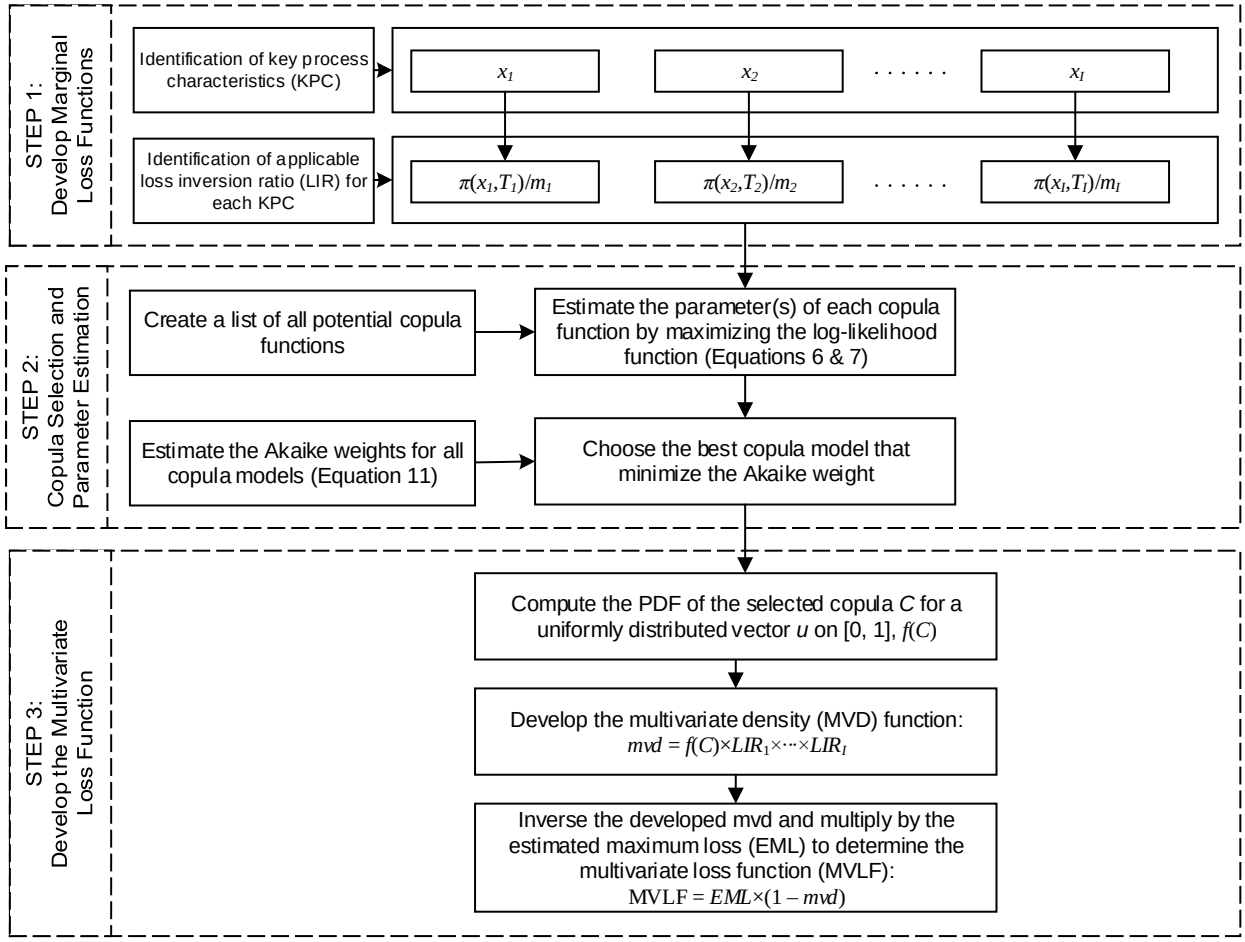


Figure 7.1. The proposed methodology for copula-based estimation of multivariate loss function

Application of copula functions is relatively straightforward using computational software packages such as R [15] and MATLAB [16]. However, the main challenges are to estimate the copula parameters and to select the best copula. A brief discussion of the copula concept is provided in Section 7.3. To overcome these challenges, methods based on maximum likelihood evaluation and information theory are then presented in Section 7.4.

7.3. Copula Functions

7.3.1. Definition

Copulas are used to describe the joint distribution of dependent random variables. With copula modelling, the marginal distributions from different families can be combined [17]. This is the main advantage of copulas compared with alternative methods, such as the use of multivariate distributions, to construct dependencies.

In this study, as shown in Figure 7.1, the copula concept is used as a mechanism to develop a joint multivariate loss density function. Considering the bivariate case, the process of developing a joint distribution function can be shown using Figure 7.2, where each pair of real numbers (x, y) leads to a point ($F(x)$, $G(y)$) in the unit square [0, 1]×[0, 1]. This mapping process, which assigns the value of the joint distribution function to each ordered pair of values of marginal distribution function is indeed a copula [18].

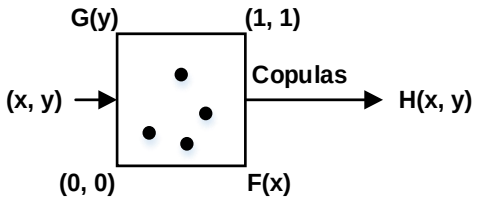


Figure 7.2. Representation of a two-dimensional (2d) copula. A 2d copula is a distribution function on a unit square [0, 1]×[0, 1], with standard uniform marginal distributions

In Figure 7.2, $F(x)$ and $G(y)$ represent cumulative distribution functions, which can be estimated using methods such as the rescaled empirical distribution function and the continuous empirical distribution function (EDF), or other estimates of the EDF including Kaplan–Meier estimate, r/n, mean rank estimate, $r/(n + 1)$, or median rank estimate, (r -

0.3)/(n + 0.4) [19], where n and r denote sample size and rank respectively (see Section 7.4.2).

To provide a formal definition for copulas, let $U_1,...,U_d$ be real random variables uniformly distributed on [0, 1]. A copula function C: $[0, 1]^d \rightarrow [0, 1]$ is a joint distribution function:

$$C(u_1,...,u_d) = P(U_1 \leq u_1,...,U_d \leq u_d).$$

Copulas are important because of the following seminal result. Let $\boldsymbol{L} = \{L_1,...,L_d\}$ be a random loss vector with continuous marginal cumulative distribution functions (CDF) with lower case letters l_i denoting assignment to loss variables. Based on Sklar's theorem [20], the CDF H of $\boldsymbol{L}$ can be represented as:

$$H(\boldsymbol{l}) = C\{F_1(l_1), \dots, F_d(l_d)\}, L \in \mathbb{R}^d,$$

in terms of a unique copula C [18]. For the proof and other important properties see [18]. The following representation of the joint density known as the copula decomposition of a joint distribution is of central interest in this work in developing the multivariate loss density function (see Step 3 in Figure 7.1):

$$H(l_1, \dots, l_d) = f_1(l_1) \times \dots \times f_d(l_d) \times c(F_1(l_1), \dots, F_d(l_d)) \qquad (7.4)$$

where c is the density of the copula [21]. The main advantage provided to the process loss analysts by this representation is that the selection of an appropriate model for the dependence between loss sources, represented by C, can then proceed independently from the choice of the marginal distributions.

7.3.2. Examples of Copulas

As described by [22], it is common to represent a copula by its distribution function. Restricting attention to the bivariate case for the sake of simplicity, copulas can be represented as:

$$C(u,v) = P(U \le u, V \le v) = \int_{-\infty}^{u}\int_{-\infty}^{v} c(x,y)dxdy, \tag{7.5}$$

where $c(x, y)$ is the density of the copula. Those copulas without a closed form solution for which the double integral at the right-hand side of Equation (7.5) are implied by well-known bivariate distribution functions, are called implicit copulas [22]. Two examples are the Gaussian copula (derived from the multivariate normal distribution) and the Student's *t*-copula (derived from multivariate Student's *t*-distribution), generally known as elliptical copulas. Compared with other copulas, the Gaussian copula has a nearly full range (-1, 1) of pairwise correlation coefficients—yielding a general and robust copula, which is used in most applications. The Gaussian copula, however, lacks the tail dependence; that is, the probability of observing extreme observations in all random variables at once. This limitation can be addressed by using Student's *t*-copula or other copula classes such as the Archimedean copula family [10]. For the case of Student's *t*-copula (or simply *t*-copula), Equation (7.5) takes the following form:

$$C_{\rho,\nu_t}(u,v) = \int_{-\infty}^{t_{\nu_t}^{-1}(u)}\int_{-\infty}^{t_{\nu_t}^{-1}(v)} \frac{1}{\sqrt{2\pi(1-\delta^2)}}\left\{1+\frac{x^2-2\rho xy+y^2}{\nu_t(1-\delta^2)}\right\}^{-(\nu_t+2)/2} dxdy,$$

where δ and ν_t are the parameters of the copula, and $t_{\nu_t}^{-1}$ is the inverse of the standard univariate *t*-distribution with ν_t degrees of freedom, expectation 0 and variance $\nu_t/(\nu_t - 2)$

[22]. A higher value for v_t decreases the probability of tail events. As the *t*-copula converges to the Gaussian copula for $v_t \to \infty$, the *t*-copula assigns more probability to tail events than the Gaussian copula [23].

On the other hand, explicit copulas are not derived from multivariate distribution functions, but do have simple closed forms. The Frank, the Clayton, and the Gumbel copulas (see Table 7.2) from the Archimedean copula family are examples of explicit copulas. Archimedean copulas are suitable for low-dimensional systems because of their simple closed functional forms. For *d*-dimensional distributions, serial iterations of Archimedean copulas are constructed, but these do not provide arbitrary pairwise correlations [10]. Other examples of this class of copulas include the Gumbel and the Clayton copulas. Table 7.2 includes the most commonly used copulas in the literature. A more comprehensive list of copulas can be found in [18] among other publications.

Table 7.2. Examples of frequently used copula functions

Copula	***C*(*u*, *v*)**	$\delta \in$
Clayton	$\left(u^{-\delta}+v^{-\delta}-1\right)^{-1/\delta}$	$\left[-1,\infty\right)/\{0\}$
Gumbel	$\exp\left\{-\left[\left(-\ln u\right)^{\delta}+\left(-\ln v\right)^{\delta}\right]\right\}^{1/\delta}$	$\left[1,\infty\right)$
Frank	$\left(-\frac{1}{\delta}\right)\cdot\ln\left\{1+\frac{\left[e^{-\delta u}-1\right]\cdot\left[e^{-\delta v}-1\right]}{e^{-\delta}-1}\right\}$	$\left(-\infty,\infty\right)/\{0\}$
Gaussian	$N_{\delta}\left(\Phi^{-1}(u),\Phi^{-1}(v)\right)$	$\left[-1,1\right]$
t	$T_{\delta,v_t}\left(T_{v_t}^{-1}(u),T_{v_t}^{-1}(v)\right)$	$\left[-1,1\right]$

As an example of how to construct a joint distribution from copulas, consider $u = F_1(x)$ and $v = F_2(y)$, where F_1 and F_2 are any two one-dimensional distributions, associated with two

random variables *X* and *Y*. Then, one joint distribution of *X* and *Y*, for example, using the Gumbel copula with a dependence parameter δ, is:

$$H(x,y) = C_\delta(F_1(x), F_2(y)) = \exp\left\{-\left[\left(-\ln F_1(x)\right)^\delta + \left(-\ln F_2(y)\right)^\delta\right]\right\}^{1/\delta}.$$

Therefore, using the copula approach, joint distributions can be constructed from arbitrary univariate distributions.

A copula is sometimes referred to as a "dependency function" since it contains all of the dependence information between random variables [24]. For instance, using the Gumbel copula, all the information about the dependence between the two random variables is in the parameter δ, with values which can be interpreted in terms of a coefficient like Kendall's τ rank correlation, because it is not affected by strictly increasing transformations of the random variables [25]. More discussion on the selection of dependence measures can be found in Hashemi, Ahmed, & Khan (2015a).

7.4. Copula Estimation and Model Selection

7.4.1. Review of the Parameter Estimation Methods

When modelling the joint density of two random variables using copula functions, care must be taken to correctly and efficiently estimate the copula parameters. Genest & Favre [12] proposed a nonparametric way of estimating the copula by using the relationship between Kendall's τ and the copula parameter to get an estimate of the latter. However, this method is suitable for explicit copulas, mainly Archimedean copulas, and requires the Kendall's τ to be known.

Methods based on maximizing the likelihood function are frequently used as an alternative method in the literature to determine the copula parameter. Two examples are the exact maximum likelihood method and the method of inference functions for margins (IFM), which estimate both the parameters of the marginals and the copula function [21]. Genest & Favre [12] proposed the canonical maximum likelihood (CML) estimation procedure that is appropriate when one does not want to specify a parametric model to describe the marginal distributions.

The efficiency and consistency of maximum likelihood estimations have been shown in several studies [12,21]. For the purpose of this study, only the parameter(s) of the copula function should be estimated since the marginal loss functions are already determined (see Figure 7.1). Therefore, the maximum log-likelihood (ML) estimator method is used in the following section to estimate the copula parameter.

7.4.2. Methodology to Estimate Copula Parameter

Let δ be the copula parameter to be estimated. Let $f_{LIRi}\left(x_i,T_i\right), i=1,...,I$ be a continuous function denoting the loss inversion ratio for key process characteristic x_i with target value T_i, and $F_{LIR_i}\left(x_i,T_i\right)$ be the LIR distribution function of X_i. Given a random sample $\left\{\left(x_{1k},...,x_{Ik}\right): k=1,...,n\right\}$ observed from distribution $F_{LIR}^{\delta}\left\{\left(x_1,T_1\right),...,\left(x_I,T_I\right)\right\} = C_{\delta}\left\{F_{LIR_1}\left(x_1,T_1\right),...,F_{LIR_I}\left(x_I,T_I\right)\right\}$, the resulting log-likelihood function for copula C with parameter δ denoted by $LL_C\left(\delta\right)$ can be represented by:

$$LL_C(\delta) = \sum_{k=1}^{n} \ln\left[c_\delta \left\{ F_{LIR_1}(x_{1k}, T_1), ..., F_{LIR_I}(x_{Ik}, T_I) \right\} \right] \tag{7.6}$$

where c_δ is the copula density. Then, the copula parameter δ is estimated using the ML estimator by maximizing the log-likelihood function of the copula density:

$$\hat{\delta} = \arg\max LL_C(\delta). \tag{7.7}$$

As discussed before, the uniform marginals $F_{LIR_i}(x_i, T_i)$ in Equation (7.6) can be estimated using rescaled versions of their empirical counterparts [12,21,27], as follows:

$$F_{LIR_i}(x_i, T_i) = \frac{1}{n+1} \sum_{k=1}^{n} 1(X_{ik} \le x_i), i = 1, ..., I. \tag{7.8}$$

The rank-based estimates, $r/(n + 1)$, can also be used to estimate the uniform marginals $F_{LIR_i}(x_i, T_i)$, where n is the sample size and r denotes the rank of each observation.

The above-mentioned ML estimation method for copula parameter estimation may seem superficially attractive both because it involves numerical work and also requires the existence of a density c_δ. At the same time, however, it is much more generally applicable than the inversion of Kendall's τ method, since it does not require the dependence parameter to be known [12]. Moreover, the application of a rank-based estimation of the distribution function $F_{LIR_i}(x_i, T_i)$ in Equation (7.6) adds to the flexibility of the approach to be applicable to the empirical loss inversion ratio $f_{LIR_i}(x_i, T_i)$ as well.

7.4.3. Review of Copula Selection Methods

Having estimated the parameters of certain copulas, another challenge is to discriminate among competing models. A goodness-of-fit test for copulas, such as that proposed by [28], and graphical methods, such as the construction of QQ-plots [27] and K-plots [12] have been used for copula selection. However, none of these methods are proven to be superior [21]. It is recommended to use different methods and compare the results, should one use the goodness-of-fit and graphical methods.

A more formal way to rank the candidate copulas is the application of Akaike's information criterion (AIC). The AIC approach, with a fundamental link to information theory, uses an empirical log-likelihood function to estimate the relative expected "information" lost, referred to as Kullback-Leibler (KL) distance, when a candidate model is used to approximate the true (real) model [29].

An alternative information-theory-based model selection method is the application of Bayesian information criterion (BIC). BIC arises from a Bayesian viewpoint with equal prior probability on each model and vague priors on the parameters, given the model. BIC is a dimension-consistent criterion that assumes that the true model remains fixed as sample size approaches infinity [29]. However, this assumption may not be applicable in most process system applications as increased sample size in process industry usually stems from the addition of new measurement sensors and inclusion of larger data sets. Accordingly, as the sample size increases, the number of factors in the model also increases. In contrast to BIC, as discussed by [29], AIC provides a scientific understanding of the process or system under study by searching for model with smallest estimated KL

distance. Therefore, AIC is used in this study as the preferred method for copula model selection. For a detailed conceptual comparison of BIC and AIC, an interested reader may refer to [29].

7.4.4. AIC Approach to Copula Selection

Accioly & Chiyoshi [19] used Equation (7.9) to calculate the AIC of each copula model through the resulting values of corresponding estimated pseudo log-likelihoods.

$$AIC = -2L(\delta) + 2P \tag{7.9}$$

where P is the number of estimable parameters. In this work, except for the *t*-copula, for other copulas in Table 7.2, $P = 1$ because the only estimable parameter is δ, given that the parameters of marginal loss functions are already determined. For the *t*-copula, $P = 2$ since in addition to δ, estimation of the degrees of freedom (v_t) is also required.

An individual AIC value, by itself, is not interpretable due to the unknown constant (interval scale). The AIC is only comparative relative to other AIC values in the model set [29]. Therefore, the best model is determined by examining their relative distance from the "true" model through computation of the AIC differences,

$$\Delta_i = AIC_i - AIC_{\min} \tag{7.10}$$

over all candidate models. The smaller Δ_i is, the more likely it is that the adjusted model is the best model. Better interpretation could also be achieved with the Akaike weights [29]:

$$w_i = \frac{\exp(-0.5\Delta_i)}{\sum_{r=1}^{R} \exp(-0.5\Delta_r)}. \tag{7.11}$$

The weight w_i is the evidence that model i is the best model, given the data and set of R candidate models. The w_i depends on the entire set; therefore, if a model is added or dropped during a post hoc analysis, the w_i must be recomputed for all the models in the newly defined set.

7.4.5. A Simulation Study

7.4.5.1. Selection of Candidate Copulas

To illustrate the application and also to evaluate the performance of the presented copula estimation and selection methodologies, a simulation study is conducted using a bivariate set of loss inversion ratios $\left[LIR(x), LIR(y)\right]$ with a presumable dependence structure. This represents two correlated inverted probability loss functions *IPLF*(x) and *IPLF*(y). The first step is to select a set of candidate copulas for this study. The Gaussian copula is selected as the traditional candidate for modelling dependence. The *t*-copula is selected as it can capture dependence in the tails of the distributions without giving up flexibility to model dependence in the center. The Gumbel copula from the Archimedean family is directly related to multivariate extensions of extreme value theory, which has gained popularity in risk management over the last few decades [23]. The Frank and the Clayton copulas from the class of Archimedean copulas are also selected due to their useful properties and ease in construction [19].

The above mentioned copulas, listed in Table 7.2, have been frequently used in the literature for modelling dependence in safety assessment [10], operational loss modelling [30], business loss modelling of process facilities [31], process loss aggregation [11], and

drilling time decision process [19]. Moreover, these copulas are included in the MATLAB® [16] statistical toolbox, which simplifies their implementation.

7.4.5.2. Parameter Estimation

For the purpose of this simulation study, data is simulated using Gaussian, *t*, Clayton, Gumbel and Frank copulas with $U(0, 1)$ marginals and parameters corresponding to Kendall's τ equal to 0.2 (low correlation) and 0.8 (high correlation). To study the effect of sample size on the performance of the parameter estimation, both 1,000 and 5,000 observations are used. According to the procedure provided in Section 7.4.2, in the first step, each pair of observations $\left[LIR(x_1), LIR(y_1)\right]$, …, $\left[LIR(x_n), LIR(y_n)\right]$ is transformed to their rank-based representation, $\left[F_{LIR}(x_i), F_{LIR}(y_i)\right]$, by

$$F_{LIR}(x_i) = \frac{rank\left(LIR_{x_i}\right)}{n+1} \text{ and } F_{LIR}(y_i) = \frac{rank\left(LIR_{y_i}\right)}{n+1}.$$

Then, Equation (7.6) is used to calculate the log-likelihood function using different copulas in Table 7.2. Finally, the parameter estimation is carried out through the maximization of the ML estimator. To assess the precision of the parameter estimation methodology, the simulation is iterated for S = 500 and S = 1,000 for each sample size. As an example, for the case of the Frank copula with $\tau = 0.2$, increasing the sample size and simulations from n = 1,000 and S = 500 to n = 5,000 and S = 1,000 resulted in about a 60% decrease in the standard deviation of the estimated parameter. The estimated mean ($\mu_{\hat{\delta}_i}$) of the copula parameters for Kendall's τ of 0.2 and 0.8 are given in Table 7.3. Table 7.3 also provides

the closed form relationships between Kendall's τ and copula parameters. These relationships are used to calculate the true copula parameter values. This simulation study shows the acceptable performance of the presented parameter estimation methodology. Moreover, as expected intuitively, higher sample size and simulation runs result in a higher precision (lower standard deviation) and higher accuracy (lower $\left|\delta_i - \mu_{\hat{\delta}_i}\right|$ value) of the estimated parameters.

Table 7.3. Estimated copula parameters for the simulation case study

			Kendall's τ = 0.2			**Kendall's τ = 0.8**		
				Estimated Parameter			**Estimated Parameter**	
Copula	**δ and τ Relationship**	**Ref.**	**True Parameter**	***n*=1000* S=500**	***n*=5000 S=1000**	**True Parameter**	***n*=1000 S=500**	***n*=5000 S=1000**
Frank	$\tau = 1 + 4\left[D_1(\delta) - 1\right]/\delta$ $D_1 = \int_0^\delta \left[t/\delta\left(e^t - 1\right)\right]dt$	[19]	1.861	1.625	1.843	18.192	19.222	17.983
Gumbel	$\tau = 1 - \delta^{-1}$	[19]	1.250	1.240	1.243	5.000	4.843	4.929
Clayton	$\tau = \delta/(\delta + 2)$	[19]	0.500	0.517	0.509	8.000	7.660	7.786
Gaussian	$\tau = (2/\pi)\arcsin(\delta)$	[12]	0.309	0.305	0.324	0.951	0.953	0.951
t	$\tau = (2/\pi)\arcsin(\delta)$	[12]	$\hat{\delta}$ = 0.309 $\hat{\theta}$ = 1.000	$\hat{\delta}$ = 0.326 $\hat{\theta}$ = 1.116	$\hat{\delta}$ = 0.316 $\hat{\theta}$ = 1.000	$\hat{\delta}$ = 0.953 $\hat{\theta}$ = 1.000	$\hat{\delta}$ = 0.953 $\hat{\theta}$ = 1.003	$\hat{\delta}$ = 0.951 $\hat{\theta}$ = 1.000

* *n*: sample size; S: number of simulations

7.4.5.3. Copula Selection

From the estimated copula parameters in Table 7.3, the associated log-likelihood function and the AIC values for each copula model with Kendall's τ of 0.2 are then calculated for both 1,000 and 5,000 sample sizes and S = 1,000 simulations. The values of AIC differences $\left(\Delta_i = AIC_i - AIC_{min}\right)$ are also calculated which allows the results to be more easily interpreted. In real-life problems, the Δ_i values can be used to rank the candidate copulas for a specific application. Experiment 1 in Table 7.4 shows the results for the case

when the Gaussian copula is used to generate the original data. As shown in Table 7.4, for n = 5,000, the copula selection methodology has selected the Gaussian copula with an Akaike weight of 0.826. The *t*-copula with a degree of freedom of 4.67×10^6 is also selected as the second best model with an Akaike weight of 0.174. The reason is that as the degrees of freedom parameter are made larger, a *t*-copula approaches the corresponding Gaussian copula [23].

Simulation study was also conducted for the other four copulas (Experiments 2 to 5 in Table 7.4). Again, the presented copula selection methodology successfully identified the original copula used to construct the dependence structure. For the sample size of n = 5,000, the differences between the original copula used (the true model) and the candidate copulas are so huge that the Akaike weights of the others can be considered as zero. Therefore, for Experiments 2 to 5, only the results for the highest Akaike weights are included to keep Table 7.4 concise. According to Table 7.4, as expected, it can be seen that the model selection procedure becomes more powerful with larger samples sizes.

Table 7.4. The copula selection results for the simulation case study for the Kendall's τ = 0.2 and sample sizes 1,000 and 5,000

Exp. No.	Copula Used	Candidate Copula	$LL_C(\delta_i)$		AIC_i		Δ_i		w_i	
			n=1000	n=5000	n=1000	n=5000	n=1000	n=5000	n=1000	n=5000
1	Gaussian									
		Gaussian	53.93	290.54	-105.85	-579.08	0.00	0.00	0.407	0.826
		t	53.92	288.98	-105.84	-575.97	0.01	3.11	0.404	0.174
		Clayton	48.56	206.26	-95.12	-410.52	10.73	168.56	0.002	0.000
		Gumbel	36.62	247.19	-71.24	-492.38	34.61	86.70	0.000	0.000
		Frank	53.14	274.93	-104.29	-547.87	1.56	31.21	0.186	0.000
2	*t*	*t*	124.59	1363.00	-247.18	-2724.00	0.00	0.00	1.000	1.000
3	Clayton	Clayton	55.36	341.71	-108.72	-681.42	0.00	0.00	0.999	1.000
4	Gumbel	Gumbel	39.72	324.79	-77.43	-647.58	0.00	0.00	0.995	1.000
5	Frank	Frank	35.94	237.75	-69.88	-473.49	0.00	0.00	0.972	1.000

A similar copula selection case study is repeated for the case of Kendall's $\tau = 0.8$. Again the results showed a good performance of the copula model selection procedure with Akaike weights of 1 for the case of sample size $n = 5{,}000$. The table of results is not included to keep the paper concise. Altogether, the AIC was found to be a good criterion for finding the best fitting copula. For a small sample size its performance may not be entirely satisfactory; however, the AIC still finds the correct model or will chose one that is close to it. The remaining question is whether an even better model might have been postulated for the models other than the candidate copula functions. Information criteria attempt only to select the best model from the candidate models available; if a better model exists, but is not offered as a candidate, then the information-theoretic approach cannot be expected to identify this new model [29]. Therefore, when using the AIC approach, it is strongly recommended to choose all possible copula functions as candidate models.

7.4.6. Copula Model Selection Uncertainty Assessment

The AIC allows a ranking of copulas and the identification of copula models that are nearly equally useful versus those that are clearly poor explanations for the data at hand. However, one must keep in mind that there is often considerable uncertainty in the selection of a particular model as the "best" approximating model. Loss function relates the process measurements to observed operational loss values. However, process measurements are noisy and consequently the shape of the loss functions, determined from the inversion of LIR distribution as a function of process measurement (see Equation 7.3), is uncertain. Thus, the observed loss values and resulting estimated LIRs are conceptualized as random

variables; their values would be different if another independent loss dataset were available. It is this "loss observation (sampling) variability" that results in uncertain statistical inference about the dependence structure from the particular loss dataset being analyzed. Various computer-intensive resampling methods may further improve the assessment of the uncertainty of inferences, but it remains important to understand that proper model selection is accompanied by a substantial amount of uncertainty.

The bootstrap technique is a type of Monte Carlo simulation which is used frequently in applied statistics for bias assessment and the evaluation of model selection uncertainty. In practical application, the empirical bootstrap means using some form of resampling with replacement from the actual data *x* to generate (e.g., 1,000) bootstrap samples. The sample data consists of *b* independent units, and it then suffices to take a simple random sample of size *n*, with a replacement, from the *b* units of data, to get one bootstrap sample. A more detailed explanation of the bootstrap method to estimate model selection uncertainty can be found in [29].

The bootstrap method is used in this section to estimate the proposed copula model selection uncertainty. Although the bootstrap method is very advantageous to allow insights into model selection uncertainty, its computer-intensive nature will continue to hinder its use for large problems. Therefore, the bootstrap sampling in this simulation study is limited to 1,000 samples due to high computation time. The Frank copula (representing a symmetrical copula) and the *t*-copula with 1 degree of freedom (representing an asymmetrical copula) are selected to generate the loss inversion ratios. To study the effect of loss sample size, both n = 1,000 and n = 5,000 pairs of $\left[LIR(x), LIR(y) \right]$ observations

are simulated using the Frank copula and the *t*-copula with $U(0, 1)$ marginals and parameter corresponding to Kendall's τ equal to 0.4 and 0.8. Then, 1,000 bootstrap samples are generated from these simulated LIR data to enable model selection uncertainty assessment due to sampling variability. Finally, for each bootstrap sample, the AIC difference and Akaike weights of a set of five candidate copula functions in Table 7.2 are computed.

Table 7.5 shows the relative model selection frequencies (π_i) from applying the AIC to each of the 1,000 bootstrap samples. From Table 7.5, it can be seen that when the Frank copula with $\tau = 0.4$ and $n = 1{,}000$ is used to simulate the original data, the model selection procedure has chosen the Frank copula as the best possible model in this group with a relatively low frequency of 0.764. However, for the case of $n = 5{,}000$ simulated LIR data, the performance of the procedure is increased significantly and the Frank copula is chosen with a relative frequency of 0.957. Table 7.5 also shows the bootstrap analysis results of using the Frank copula with $\tau = 0.8$ where both $n = 1{,}000$ and $n = 5{,}000$ simulated LIR data and the model selection procedure selected the Frank copula with a relative frequency of 1. Similar performance is also observed for the case that the *t*-copula is used to generate the LIR data.

From the bootstrap analysis, it can be concluded that the uncertainty of the model selection procedure decreases significantly with increasing sample size. Moreover, the copula selection uncertainty is lower for higher rank correlation values as it is intuitively easier to recognize the dependence structure for highly correlated data. An important conclusion is that using a large sample size is crucial to study the dependence structure among variables and select the correct copula, especially for low correlation values.

Table 7.5 also shows the calculated Akaike weights. It can be seen that the relative frequencies for copula *i* being selected as the best copula are similar to the Akaike weights, but are not identical. This important observation shows that there is not any particular advantage in the bootstrap selection frequencies over the Akaike weights. In other words, comparing the results in Table 7.5, one can assess the model selection uncertainty directly from Akaike weights. Considering the extensive computation time required for the computation of the model selection frequencies using bootstrap analysis, the Akaike weights, in general, can be used as a preferred method for both model selection and uncertainty assessment.

Moreover, as shown in Table 7.5, compared to the bootstrap analysis, considerably smaller sample sizes (n = 500 and n = 1,000) have provided a good support for the best copula using Akaike weights. This means that using Akaike weights for uncertainty assessment of copula selection is much less sensitive to the loss observations sample size than the bootstrap analysis. Overall, Akaike weights are easy to estimate with minimal computation time and provide a better estimate of the best model, given an a priori set of models. This conclusion is also consistent with the observations by Burnham & Anderson [29] on the advantages of Akaike weights.

Table 7.5. Comparison of bootstrap and Akaike weight performance in model selection uncertainty analysis

Copula Used	Bootstrap Selection Frequency (π_i)				Akaike Weight (w_i)			
	τ = 0.4		τ = 0.8		τ = 0.4		τ = 0.8	
	n=1000	n=5000	n=1000	n=5000	n=500	n=1000	n=500	n=1000
Frank	0.764	0.957	1.000	1.000	0.95	1.00	1.00	1.00
t	0.806	0.998	0.996	1.000	0.98	1.00	1.00	1.00

7.5. Case Study: Separation Column

7.5.1. Case Study Description

The practical application of the proposed multivariate operational loss model is demonstrated using a de-ethanizer column case study. The de-ethanizer simplified process schematic and the feed and product characteristics are depicted in Figure 7.3. A liquid feed stream, consisting of a mixture of hydrocarbon components to be separated, is fed into the column. If the top product (ethane) is within specification (≤3% C_3), it is fed to a downstream unit for further processing and transportation to the market. Off-specification ethane goes to a tank and may be reprocessed or used as fuel (which is of lower value). Similarly, the bottom product (C_{3+}) is used in another part of the plant or fed to a pipeline to produce a higher-value product if it meets specifications (≤5% C_2), and any off-specification product may be sent to a tank for reprocessing.

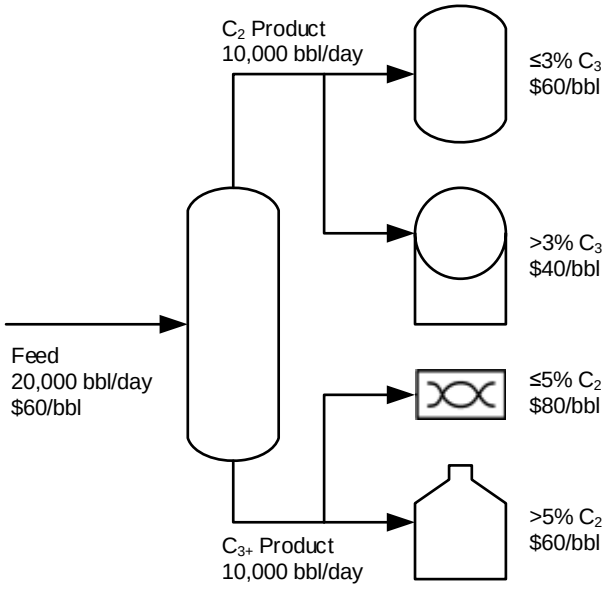


Figure 7.3. Feed and product characteristics for the de-ethanizer column case study

7.5.2. Economic Analysis of Separation

Flooding is a common abnormal process condition that can cause loss of separation and negatively impact the safety and energy efficiency of the separation process. Different measurable process variables can be used as key process characteristics (KPC) to indicate the flooding conditions in a separation column. For a typical de-ethanizer, a simultaneous increase in the column bottom temperature (or reboiler temperature) derivative and the differential pressure derivative across the column could be an indication of flooding conditions [32]. Accordingly, the de-ethanizer column bottom temperature (T_c) and differential pressure (*DP*) are considered as the KPCs in this case study. Under normal operating conditions, these two KPCs are considered to be normally distributed with the values of mean and standard deviation as follows: $\mu_{DP} = 51$ (set point) and $\sigma_{DP} = 3.75$ in millimetre (mm) of H_2O, $\mu_{T_c} = 82$ (set point) and $\sigma_{T_c} = 5$ in degrees centigrade (^{o}C). Using the proposed methodology, it is easy to expand the study from bivariate to multivariate analysis by considering more monitored process variables.

The operational loss due to the variability of the separation process includes the cost of increased energy usage and decreased product value. As noted in Figure 7.3, the off-specification products have lower value compared to on-specification products. The operational loss data in Table 7.6 are considered for the operation of the de-ethanizer column. The financial information and process characteristics used in this case study are for illustrative purposes. This information is used in the following section to determine the loss functions shape parameter.

As shown in Table 7.6, the operational loss is divided into Class 0-allowable operational loss and Class I-unallowable operational loss. The Class 0 is the loss due to normal fluctuation of the process characteristic(s) within the specification limits. The Class 0 is referred to as allowable operational loss as organizations accept that people, processes and systems are imperfect and that losses will arise from errors and ineffective operations. The Class I loss is the unallowable operational loss recorded as a result of deviation of process characteristic(s) beyond the specification limits. The subsequent unprofitable process operation may result in production of sub-quality products, increased energy usage, and unsafe process conditions.

Table 7.6. The assumed operational loss information for the de-ethanizer column

KPC	Symbol	Value	Alarm Tag	Associated Operational Loss ($/bbl)	Description
Differential Pressure (mm H_2O)	DP_T	51	-	0	Target differential pressure
	$DP_T + 3\sigma_{DP}$	62.25	Hi	4.2	Maximum Class 0 operational loss for column overpressure.
	$DP_{\max}$	84	Hi-Hi	20	Maximum Class I operational loss for column overpressure.
	$DP_T - 3\sigma_{DP}$	39.75	Low	6.5	Maximum Class 0 operational loss for column under-pressure.
	$DP_{\min}$	34	Low-Low	12	Maximum Class I operational loss for column under-pressure. A lower value of operational loss is considered for under-pressure of the column, compared to overpressure case, due to different impact on the separation process.
Column Bottom Temperature (^{o} C)	T_{cT}	82	-	0	Target column bottom temperature
	$T_{cT} + 3\sigma_{T_c}$	97	Hi	4	Maximum Class 0 operational loss for column over-temperature.
	$T_{c\max}$	135	Hi-Hi	20	Maximum Class I operational loss for column over-temperature.
	$T_{cT} - 3\sigma_{T_c}$	67	Low	4.2	Maximum Class 0 operational loss for column under-temperature.
	$T_{c\min}$	55	Low-Low	14	Maximum Class I operational loss for column under-temperature.

7.5.3. Description of Incident Scenarios

A hypothetical overpressure scenario in the de-ethanizer column is considered, where the failure of the existing control systems resulted in flooding conditions. Eventually, the column DP exceeded a critical threshold and triggered flooding. Once the high-high alarm triggered, the operator successfully diagnosed and corrected the process fault by cutting the reboiler heat duty. The flooding conditions began relaxing during the reboiler's interrupted period. Consequently, after a few minutes the column bottom temperature and

differential pressure set points returned to target values. No process downtime or loss of containment happened; however, the production of sub-quality products during flooding conditions resulted in operational loss.

Similarly, a column under-temperature/under-pressure scenario is also considered in this case study as another cause of operational loss due to the production of sub-quality products.

7.5.4. Marginal Loss Functions

As mentioned in Figure 7.1, the first step in implementation of the proposed multivariate operational loss methodology is to assign a loss function for each identified KPC. In an earlier study, Hashemi et al. [3] concluded that modified INLF (MINLF) and IBLF demonstrate better performance than other IPLFs as their shape can be modified more flexibly to suit the practitioner's needs for both symmetric and asymmetric problems. Using the loss data points identified in Table 7.6 and the search algorithm based on the least-squares method in [3], the shape parameters of the MINLF and IBLF for column over-temperature and under-temperature cases are determined as shown in Figure 7.4. The loss values at low alarm ($T_{cT} - 3\sigma_{T_c}$) and high alarm ($T_{cT} + 3\sigma_{T_c}$) set points for column bottom temperature used to determine the loss functions shape parameters are also shown in Figure 7.4. From Figure 7.4, it can be seen that IBLF fits the system loss behaviour better than the MINLF.

Similarly, Figure 7.5 illustrates the resulting MINLF and IBLF and associated shape parameters for column overpressure and under-pressure cases along with the loss values at

low and high alarms set points for column pressure. From Figure 7.5, it can be seen that both IBLF and MINLF demonstrate almost similar performances for column DP. As IBLF is selected for the column bottom temperature, it is decided to choose MINLF for column DP to show the flexibility of the proposed multivariate loss function methodology in using different marginal loss functions.

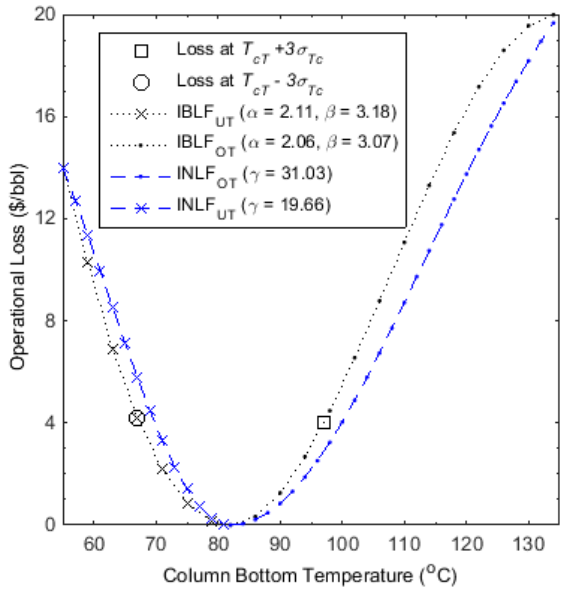


Figure 7.4. MINLF and IBLF and associated shape parameters for de-ethanizer column bottom temperature (UT: under-temperature; OT: over-temperature)

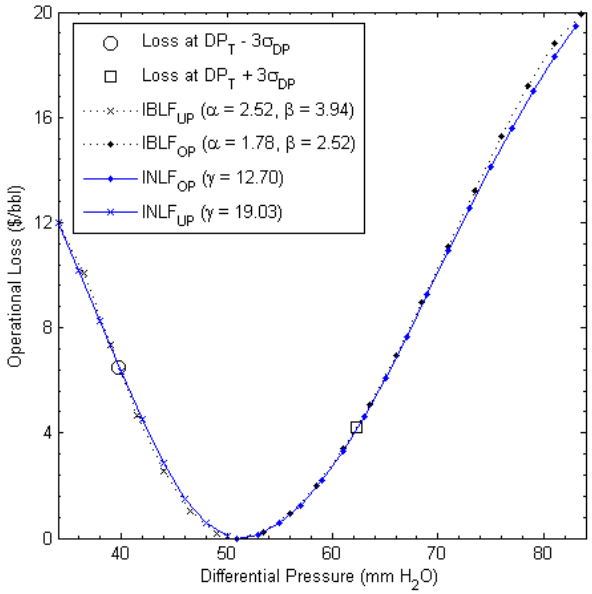


Figure 7.5. MINLF and IBLF and associated shape parameters for de-ethanizer column differential pressure (UP: under-pressure; OP: overpressure)

7.5.5. Development of Multivariate Loss Function

Proceeding with the multivariate analysis, the copula parameter estimation and copula selection procedures proposed in Section 7.4 are used to determine the best copula for construction of the multivariate loss function. Table 7.7 represents maximum likelihood estimation results and the estimated parameters together with AIC differences and Akaike weights for different copulas. As can be seen, there is a good support to show that the *t*-copula model is the best possible model in this set of models since its Akaike weight is significantly greater than the others. Figure 7.6(a) shows the contour plot and Figure 7.6(b) shows the three-dimensional plot of the multivariate loss function developed using *t*-copula for the de-ethanizer column, following Step 3 of the methodology in Figure 7.1.

From Figure 7.6, one can see that the de-ethanizer process experiences zero operational loss when both column bottom temperature and differential pressure are on target (i.e. DP_T = 51 mm H_2O and T_{cT} = 82^o C). For the overpressure and over-temperature (OP-OT) scenario, the system's operational loss attains its maximum value at \$20/bbl (see Figure 7.6). For the under-pressure and under-temperature (UP-UT) scenario, the system attains lower maximum loss values as indicated in Table 7.6. Comparing the developed multivariate loss function with the loss information provided in Table 7.6, it can be seen that the proposed multivariate loss function approach has a good performance in modelling the system loss when dealing with more than one key process characteristic.

Table 7.7. Maximum likelihood estimation results with calculated AIC differences and Akaike weights for the de-ethanizer column

Copula	$\hat{\delta}_i$	$L(\delta_i)$	AIC_i	Δ_i	w_i
Gaussian	0.02	-0.04	2.08	4.30	0.08
t	0.01	3.11	-2.21	0.00	0.65
Clayton	0.02	0.35	1.30	3.51	0.11
Gumbel	1.02	0.09	1.83	4.04	0.09
Frank	0.04	0.01	1.99	4.20	0.08

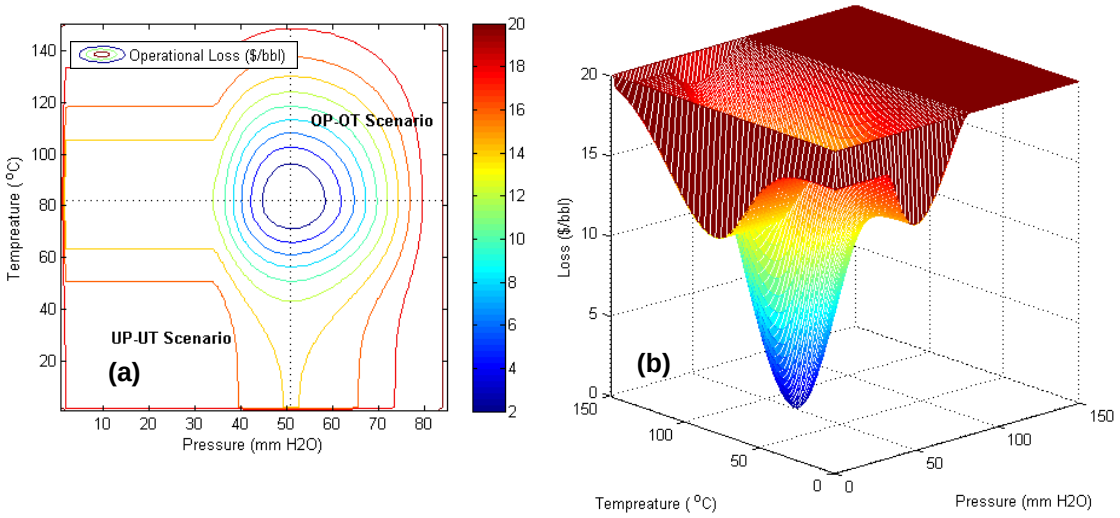


Figure 7.6. (a) Contour plot and (b) three-dimensional plot of the multivariate loss function developed using the *t*-copula for the de-ethanizer column

From Table 7.7, it can be concluded that the Clayton and the Gumbel copulas from the Archimedean family are the next two copulas which can be used for this case study. It is not surprising to see that the first three copulas with the lower Akaike differences in Table 7.7 are of the extreme-value type since the marginal loss functions are asymmetric. The Frank and Gaussian copulas, both with a radial symmetry property, have the largest Akaike differences. As mentioned before, the larger Δ_i is, the less plausible it is that the candidate model is the best model. Table 7.8 provides some rules of thumb that can be used to identify the level of support of different candidate models. Based on the guidelines in Table 7.8 and the calculated Akaike differences in Table 7.7, it can be concluded that four other copulas other than the *t*-copula can also be taken into consideration as an approximating dependence model for further analysis. The multivariate loss functions determined using Clayton, Gumbel and Frank copulas are shown in Figure 7.7(a-c). Comparison of Figures 7.6(a) and 7.7(a-c) indicates that all of these copula functions considered in this case study are able to provide the overall dependence structure between the column differential pressure and bottom temperature. However, as concluded from the calculated Akaike weights in Table 7.7, for this study, *t*-copula provides the best approximating model.

Table 7.8. AIC differences and level of support for candidate models [29]

Δ_i	**Level of Empirical Support for Model *i***
0 - 2	Substantial
4 - 7	Considerably less
> 10	Essentially none

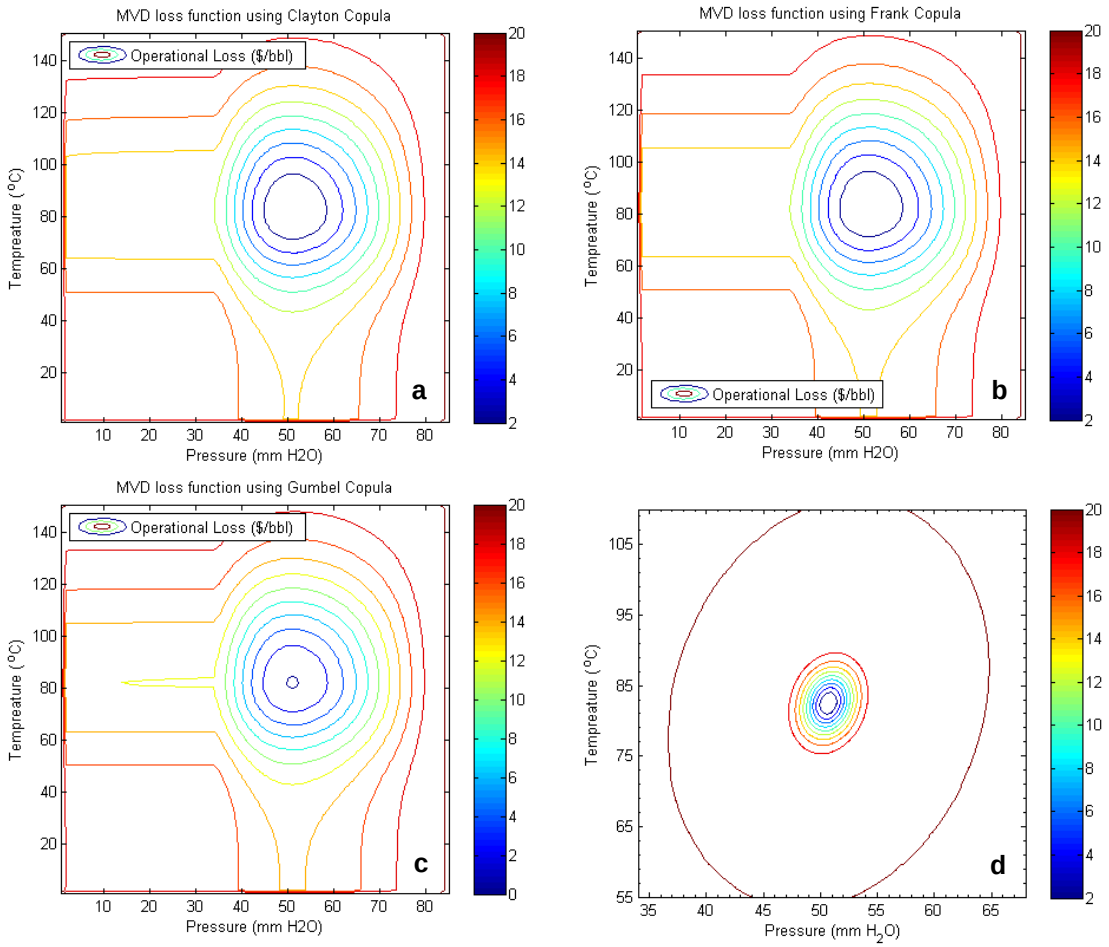


Figure 7.7. Contour plots of the multivariate loss function developed using (a) Clayton copula, (b) Frank copula, (c) Gumbel copula and (d) bivariate INLF (Equation 7.1)

7.5.6. Discussion

To compare the performance of the proposed copula-based multivariate loss function with traditional approaches, Equation (7.1) is used to develop a bivariate loss function based on the inverted bivariate normal loss function. For this case study, Equation (7.1) takes the following form:

$$L(\boldsymbol{Y}) = EML\left[1-\exp\left\{-\frac{1}{2}\left(\begin{bmatrix} DP \\ T_c \end{bmatrix}-\boldsymbol{T}\right)^{\mathrm{T}}\cdot\begin{pmatrix}\gamma_1 & \gamma_{12}\\ \gamma_{12} & \gamma_2\end{pmatrix}^{-1}\cdot\left(\begin{bmatrix} DP \\ T_c \end{bmatrix}-\boldsymbol{T}\right)\right\}\right]$$

where $\boldsymbol{T} = \begin{bmatrix} DP_t \\ T_{cT} \end{bmatrix}$ and DP_t, T_{cT} and *EML* are to be replaced from Table 7.6. In the equation above, $\gamma_1 = \Delta_{DP}/4$, Δ_{DP} being the Euclidean distance from $\boldsymbol{T}$ to the point where maximum loss occurs along the principal (*DP*) axis; $\gamma_2 = \Delta_{Tc}/4$, Δ_{Tc} being the Euclidean distance from $\boldsymbol{T}$ to the point where maximum loss occurs along the secondary (T_c) axis; and $\gamma_{12} = \eta\Delta_{T_c}^2/16$, where η is the slope of the principal axis [7]. From the structure of equation above, the following limitations are identified:

i. Expanding this method to multivariate systems is difficult due to mathematical restrictions.
ii. Only INLF should be used for both marginal loss functions.
iii. Only symmetrical loss functions can be used for marginal loss functions. In other words, for instance, the maximum loss for both over and under-pressure scenarios should be the same, which may not be the case for real-life applications.
iv. There is no straightforward method to estimate η.

Figure 7.7(d) shows the bivariate loss function for the de-ethanizer column using the approach described above, where $\eta = 0.01$ is found as the best estimate to represent the loss information in Table 7.6. Comparing the contour plots in Figure 7.7(d) with the loss information provided in Table 7.6, significant differences between the estimated loss values and the actual loss information can be seen. The bivariate loss function shown in Figure 7.7(d) overestimates the operational loss as the system almost attains its maximum loss even when *DP* and T_c are within the specification limits.

The proposed copula-based multivariate loss function approach overcomes the limitations of the traditional approaches based on classical multivariate distributions and provides a more accurate and realistic estimation of the system's operational loss. As demonstrated using this case study, the proposed copula-based methodology has the following advantages:

- The method can be easily expanded from a bivariate case to a multivariate case.
- Any type of symmetrical or asymmetrical loss function can be used for the marginal loss functions.
- Estimation of the marginal loss functions and the dependence structure can be done separately, which simplifies the practical application of the methodology.
- The simulation study in Section 7.4 demonstrates the efficiency of the proposed framework in estimating the copula parameter and choosing the best copula.

7.6. Conclusions

In this paper, a methodology to construct the multivariate loss functions is proposed using copula functions, which allows selection of any type of inverted probability loss function for the marginal loss functions irrespective of their dependence structure. Although application of copula functions in practical problems is straightforward using the existing computational software, challenges exist in estimation of the copula parameter(s) and selection of the best copulas. To address these challenges, a method based on maximum likelihood evaluation and the Akaike information criterion (AIC) is presented and its efficacy is demonstrated using a simulation case study. The simulation study showed that

the performance of copula estimation procedure is acceptable for both low and high correlations with more satisfactory results for larger sample sizes. The AIC approach heavily depends on the set of selected copula models; thus, a broad set of candidate copulas should always be favoured. Bootstrap analysis is used to assess the uncertainty of the copula selection procedure due to the sampling variability and the results are found to be consistent with the copula ranking based on Akaike weights. Therefore, it is concluded that Akaike weights, in general, can be used as a preferred method for both model selection and uncertainty assessment.

The overall methodology is applied to a separation case study. From the case study results, it is observed that several copulas provide acceptable models of the dependence in the separation column under study. The copulas with the highest Akaike weights are of the extreme-value type and a comparison study showed that all candidate models have some level of support for modelling purposes. For this case study, the *t*-copula is identified as the best approximating model by having the largest Akaike weight. Although the case study presentation was limited to the case of two variables, the methodology described here extends to the multidimensional case. As the number of variables increases, of course, the intricacies of modeling become more complex. The case study results show a significantly improved representation of the process loss behaviour when using the presented copula-based multivariate loss function methodology instead of the classical multivariate inverted normal loss function approach. Even for independent losses (or any random variables), the copula approach is still a useful tool for constructing multivariate distributions, where the

advantage is solely due to the flexibility in using any form of marginal univariate distribution.

The combination of the multivariate operational loss modeling proposed in this work with a probabilistic approach leads to a multivariate methodology for operational risk assessment of process industries. To achieve this goal, further research is required to apply the copula approach in probability assessment of abnormal situations for multivariate processes.

7.7. References

[1] Leung BPK, Spiring FA. Some Properties of the Family of Inverted Probability Loss Functions. Qual Technol Quant Manag 2004;1:125–47.

[2] Tahera K, Ibrahim RN, Lochert PB. The effect of non-constant process variance in determining the optimal process means and production run of a deteriorating process. Prod Plan Control 2010;21:36–46. doi:10.1080/09537280903239433.

[3] Hashemi SJ, Ahmed S, Khan F. Loss functions and their applications in process safety assessment. Process Saf Prog 2014;33:285–91. doi:10.1002/prs.11659.

[4] Hashemi SJ, Ahmed S, Khan F. Risk-based operational performance analysis using loss functions. Chem Eng Sci 2014;116:99–108. doi:10.1016/j.ces.2014.04.042.

[5] Chan WM, Ibrahim RN. Evaluating the quality level of a product with multiple quality characteristics. Int J Adv Manuf Technol 2004;24:738–42. doi:10.1007/s00170-003-1751-6.

[6] Pignatiello JJ. Strategies for robust multiresponse quality engineering. IIE Trans 1993;25:5–15. doi:10.1080/07408179308964286.

[7] Spiring FA. The reflected normal loss function. Can J Stat 1993;21:321–30.

[8] Drain D, Gough AM. Applications of the upside-down normal loss function. IEEE Trans Semicond Manuf 1996;9:143–5.

[9] Meel A, Seider WD. Real-time risk analysis of safety systems. Comput Chem Eng 2008;32:827–40. doi:10.1016/j.compchemeng.2007.03.006.

[10] Pariyani A, Seider WD, Oktem UG, Soroush M. Dynamic Risk Analysis Using Alarm Databases to Improve Process Safety and Product Quality: Part II - Bayesian Analysis. AIChE J 2012;58:826–41. doi:10.1002/aic.12642.

[11] Hashemi SJ, Ahmed S, Khan FI. Loss scenario analysis and loss aggregation for process facilities. Chem Eng Sci 2015;128:119–29. doi:10.1016/j.ces.2015.01.061.

[12] Genest C, Favre A-C. Everything You Always Wanted to Know about Copula Modeling but Were

Afraid to Ask. J Hydrol Eng 2007:347–68.

[13] Sun F-B, Laramee J-Y, Ramber JS. On Spiring's normal loss function. Can J Stat 1996;24:241–9.

[14] Leung BPK, Spiring FA. The inverted beta loss function: properties and applications. IIE Trans 2002;34:1101–9.

[15] R Core Team. R: A language and environment for statistical computing 2013.

[16] MathWorks. MATLAB Release 2013a 2013.

[17] Genest C, Neslehová J. A Primer on Copulas for Count Data. Astin Bull 2007;37:475–515. doi:10.2143/AST.37.2.2024077.

[18] Nelsen RB. An Introduction to Copulas. 2nd ed. New-York: Springer–Verlag; 2006.

[19] Accioly R, Chiyoshi FY. Modeling dependence with copulas: a useful tool for field development decision process. J Pet Sci Eng 2004;44:83–91. doi:10.1016/j.petrol.2004.02.007.

[20] Sklar A. Fonctions de répartition à n dimensions et leurs marges. Publ Inst Stat Univ Paris 1959:229–31.

[21] Manner H. Estimation and Model Selection of Copulas with an Application to Exchange Rates. Maastricht, Netherlands: 2007.

[22] Aas K. Modelling the dependence structure of financial assets: A survey of four copulas. Oslo, Norway: 2004.

[23] Kole E, Koedijk K, Verbeek M. Selecting copulas for risk management. J Bank Financ 2007;31:2405–23. doi:10.1016/j.jbankfin.2006.09.010.

[24] Klaus M. Multivariate Dependence Modeling using Copulas (Master's thesis). Charles University in Prague, 2012.

[25] Andrade DF, Barbetta PA, De Freitas Filho P, De Mello Zunino N, C. Jacinto C, Filho PJDF. Using Copulas in Risk Analysis. Proc. 2006 Winter Simul. Conf., Brazil: IEEE; 2006, p. 727–32. doi:10.1109/WSC.2006.323152.

[26] Hashemi SJ, Ahmed S, Khan F. Correlation and Dependency in Multivariate Process Risk Assessment. IFAC SAFEPROCESS 2015 9th IFAC Symp. Fault Detect. Superv. Saf. Tech. Process. IFAC-PapersOnLine, vol. 48, Paris: Elsevier Ltd.; 2015, p. 1339–44. doi:10.1016/j.ifacol.2015.09.711.

[27] Accioly R, Aiube FAL. Analysis of Crude Oil and Gasoline Prices through Copulas. Cad Do IME - Ser Estatística 2008;24:15–28.

[28] Fermanian J-D. Goodness of fit tests for copulas. J Multivar Anal 2005;95:119–52.

[29] Burnham KP, Anderson DR. Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach. 2nd ed. Secaucus, NJ, USA: Springer; 2002.

[30] Soprano A, Crielaard B, Piancenza F, Ruspantini D. Measuring Operational and Reputational Risk – A Practitioner's Approach. Hoboken, NJ.: Wiley; 2010.

[31] Hashemi SJ, Ahmed S, Khan F. Probabilistic modelling of business interruption and reputational losses for process facilities. Process Saf Prog 2015;34:373–82. doi:10.1002/prs.11753.

[32] Dzyacky GE. Process and apparatus for predicting and controlling flood and caryover conditions in a separation column. 5784538, 1998.

8. MULTIVARIATE PROBABILISTIC SAFETY ANALYSIS OF PROCESS FACILITIES USING THE COPULA BAYESIAN NETWORK MODEL[7]

Preface

A version of this manuscript is accepted for publication in the Journal of Computers & Chemical Engineering. I am the primary author of this paper. Along with the co-authors, Faisal Khan and Salim Ahmed, I developed the conceptual model and subsequently translated this to the numerical model. I conducted the literature review and proposed an alternate modelling approach to address the shortcomings of the Bayesian network analysis. I carried out most of the data collection and the comparison of loss functions. I prepared the first draft of the manuscript and subsequently revised the manuscript based on the co-authors' feedback and also the initial feedback from the journal reviewers. The co-author Faisal Khan helped in developing and testing the concepts/models, reviewed and corrected the models and results, and contributed in preparing, reviewing and revising the manuscript. The co-author Salim Ahmed contributed through support in the development, testing and improvement of the models. Salim Ahmed also assisted in reviewing and revising the manuscript.

[7] Hashemi et al. Computers & Chemical Engineering, 93 (2016) 128–142.

Abstract

Integrated safety analysis of hazardous process facilities calls for an understanding of both stochastic and topological dependencies, going beyond traditional Bayesian Network (BN) analysis to study cause-effect relationships among major risk factors. This paper presents a novel model based on the Copula Bayesian Network (CBN) for multivariate safety analysis of process systems. The innovation of the proposed CBN model is in integrating the advantage of copula functions in modelling complex dependence structures with the cause-effect relationship reasoning of process variables using BNs. This offers a great flexibility in probabilistic analysis of individual risk factors while considering their uncertainty and stochastic dependence. Methods based on maximum likelihood evaluation and information theory are presented to learn the structure of CBN models. The superior performance of the CBN model and its advantages compared to traditional BN models are demonstrated by application to an offshore managed pressure drilling case study.

Key words: Correlation; dependence structure; multivariate probabilistic model; Akaike's information criterion.

8.1. Introduction

Process safety and risk assessment are often multidimensional and hence require the study of several potentially correlated random variables from different risk sources. Consequently, risk practitioners usually deal with complex process systems with multiple correlated variables rather than considering independent risk factors. Looking for relationships among variables is an essential part of process safety analysis to understand

the system, identify the cause(s) of process symptoms, predict abnormal conditions and protect systems from catastrophic events. There are some recent works extracting and analyzing interrelationships among random variables in the context of process facilities [1–3]. However, development of a tool to simultaneously capture different aspects of variables' interrelationships (including causality and dependency) in complex systems with high dimensionality is still a formidable challenge.

In recent years, Bayesian Network (BN) analysis has been used in process safety analysis mainly through multivariate probabilistic analysis [2] and probability updating [4]. However, BN analysis has some restrictions from the multivariate analysis perspective, which are mainly lack of control of the marginal distribution of variables and inability to capture the non-linear dependence structure [2].

To address the limiting properties of BN analysis, Elidan [5] proposed the Copula Bayesian Network (CBN) that fuses the frameworks of the statistical copula and BNs. Copulas allow the modelling of complex real-valued distributions by separating the choice of the univariate marginal distributions and the dependence function that "couples" them into a coherent joint distribution [6]. In contrast to BN analysis that uses conditional probability distributions to define a joint density, in the CBN model a collection of local copula functions is used to capture the direct dependence among system variables [5].

The objective of this work is to address the limitations of traditional BN analysis by adopting the concept of the CBN model for application in multivariate probabilistic analysis of abnormal conditions in process facilities. The contribution of this work is twofold. Firstly, by using the language of probabilistic graphical models, this work applies

copula functions to extend the BN applications in the context of process facilities' safety analysis of higher dimensions. Secondly, a learning mechanism based on a combination of maximum likelihood evaluation and information theory is introduced to address the issue of structure learning for CBN models.

The rest of the paper is organized as follows. In Section 8.2 a comparison of the complementing properties of BN analysis and copula functions is provided, followed by a discussion of the interesting synergy that can be achieved by combining copulas and BNs. The proposed CBN model for process safety analysis is provided in Section 8.3. The practical application of the proposed model is demonstrated using a case study in Section 8.4, followed by some concluding remarks.

8.2. Preliminaries

8.2.1. Inter-Relationships of Process Variables

Multivariate probabilistic process safety analysis requires identification of the inter-relationships among process variables. Connectivity, causality, and correlation are three important attributes which are used to describe such inter-relationships [7]. The illustrative example in Figure 8.1, adopted from Yang et al. [7], is provided to facilitate better explanation of the physical interpretation and practical use of these different concepts. As shown in Figure 8.1, a liquid hydrocarbon feed stream is fed into a distillation column. Following the principle of fluid dynamics, the feed flow rate (F_1) influences the liquid level in the column (L) and L influences the bottom product flow rate (F_3). In terms of the information flow path, the signal line is connected to valve V_2 to transmit the level signal

L to the control valve. This connectivity is shown in Figure 8.1(b) However, F_3 also influences L, which is different from connectivity. In fact, valve V_2 controls the flow rate F_3 based on the signals transmitted from the level meter to the control valve. The same causality relationship also exists for the overhead flow rate (F_2) and the column top pressure P (Figure 8.1.b). Thus, causality does not exist without connectivity.

To describe the concept of correlation, consider the flow rate F_1 which affects both flow rates F_2 and F_3. Intuitively, F_2 and F_3 are correlated, which can be shown by investigating the process data. However, there is no causality between F_2 and F_3, since by ruling out their common cause, F_1, F_2 and F_3 are both independent.

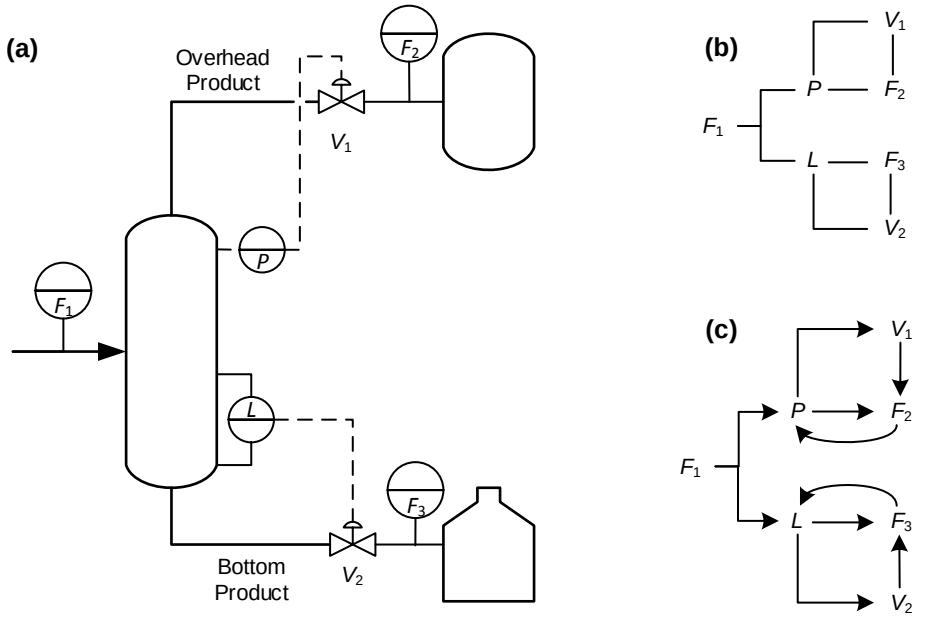


Figure 8.1. Distillation column example. (a) Schematic; (b) Connectivity; (c) Causality; adopted from Yang et al. [7]

From the illustration above it can be concluded that correlation is a necessary (not sufficient) condition for causality. Beside correlation, some additional conditions are required to imply a causal relationship, such as connectivity, responsiveness, and the direction of the relationship between two process variables [7]. Therefore, different tools are required to capture causality and dependency among variables.

Process knowledge can be used to capture causality using common methods such as structural equation models (SEM), graphical models, and rule–based models [7]. However, as reliable process knowledge is not always available, it is also important to explore capturing causality from process data. Lag-based methods, such as Granger causality and transfer entropy, and conditional independence methods, such as BNs, are widely used approaches to capture causality from process data. Linear relationships among process variables and stationary data time series are restrictive assumptions of Granger causality and transfer entropy, respectively [7]. Application of BN analysis is of main interest in this work to capture causality due to its ability to represent intuitive cause-effect relationships among process variables, as well as several other modelling advantages as discussed in Section 8.2.2.

To measure the dependency, the common approach in process facilities literature has been the application of correlation coefficients. The Pearson ρ for linear relationships and rank correlation coefficients (such as Spearman's ρ_s and Kendall's τ) for nonlinear relationships are the frequently used correlation coefficients [8]. However, given the complexity of relationships among process variables, dependence can be quantified in more sophisticated ways than merely through these numeric coefficients. Copula functions, sometimes

referred to as “dependency functions”, contain all of the dependence information among random variables [9]. Using copulas, the dependence pattern of the random variables and their individual behaviours (more precisely, their marginal probability distributions) can be studied separately.

Beyond modelling dependency and causality, copulas and BNs are both widely used in the literature to provide a framework for modelling multivariate distributions. In the subsequent subsections, a brief review of advantages and shortcomings of both approaches is provided first. Then, the potential synergy from the integration of copulas and BNs to allow simultaneous modeling of stochastic and topological dependencies among process variables is discussed.

8.2.2. Bayesian Networks (BNs)

BN analysis offers a general framework for analyzing causal influences and constructing multivariate distributions. Basically, BNs are probabilistic networks which rely on Bayes’ theorem to draw inferences based on prior evidence [10]. A BN can be defined as a directed acyclic graph (DAG) associated with a joint probability distribution [11]. BNs’ main application in process safety analysis is as an inference engine for updating the prior occurrence probability of events given new information [12,13]. This advantage addresses one of the main shortcomings of the traditional fault tree, event tree, and bowtie safety analysis methods. However, despite the broad scope of applicability, the following shortcomings are identified for BN applications:

i. Deterministic point-based probability values are used in most BN applications, ignoring the uncertainty associated with probability estimations.

ii. To tackle the above shortcoming, Gaussian distribution has been used as the marginal distribution in some applications. However, there is no doubt that the assumption of joint normality fails to yield suitable models in many applications. Aside from the case of the normal distribution, application of other probability distributions for marginal distributions is not practical due to the limitations of the BN structure [5].

iii. Constructing conditional probability tables (CPTs) to describe the strength relationships quickly becomes very complex and difficult to compute as the number of parents and states increases [2].

iv. Furthermore, representation of the dependence structure is simply limited to the definition of nodes' relationships using CPTs. Therefore, BN models fail to model complex non-linear dependencies.

There have been some recent developments to improve the practical application of BNs, such as the application of multinomial likelihood functions [14] and nonlinear Gaussian belief networks [15] to model non-linear interactions, and the application of object-oriented BN [16] and Noisy-OR technique [17] to simplify the analysis of complex networks. Although such developments have enhanced the practical implementation of the BNs, the limitations mentioned above still exist.

8.2.3. Copulas

An alternate and markedly different approach for constructing multivariate distributions is the application of copula functions to link univariate marginal distributions. Let $\boldsymbol{U} = (U_i)$, $i \in \{1, ..., d\}$ and $d \in N$ be real random variables marginally uniformly distributed on [0, 1]. A copula function $C:[0,1]^d \rightarrow [0,1]$ is a joint distribution function:

$$C_\delta\left(u_1,...,u_d\right) = P\left(U_1 \le u_1,...,U_d \le u_d\right).$$

where δ is the parameter of the copula function. The importance of copulas is rooted in Sklar's theorem [18] that states any multivariate distribution can be represented as a copula function of its marginal [6]. Given a family $\boldsymbol{X} = (X_i)$, $i \in \{1, ..., d\}$ and $d \in N$, of continuous random variables, this relationship can be stated in terms of probability density function (PDF) using the derivative chain rule:

$$f\left(\boldsymbol{x}\right) = \frac{\partial^d C_\theta\left(F_1\left(x_1\right),...,F_d\left(x_d\right)\right)}{\partial F_1\left(x_1\right)...\partial F_d\left(x_d\right)} \prod_{i=1}^{d} f_i\left(x_i\right) \equiv c_\theta\left(F_1\left(x_1\right),...,F_d\left(x_d\right)\right) \prod_{i=1}^{d} f_i\left(x_i\right), \qquad (8.1)$$

where the copula density c is uniquely determined for continuous random variables [5]. For the proof of Equation (8.1) and other important copulas and their properties, see Nelsen [6].

In practice, copula constructions often lead to a significant improvement in density estimation. Accordingly, there has been a growing interest in application of copulas in the process industry [19], with applications ranging from multivariate loss modelling and loss aggregation for process facilities [1,20,21] to risk analysis of safety systems [13,22].

A recent contribution to construct joint probability distributions using copulas is the "rolling pin method" proposed by Mohseni Ahooyi et al. [22] to accommodate random

variables with arbitrary (nonmonotonic or monotonic) relationships. However, only symmetrical copulas can be used in this method to approximate dependence structures. Moreover, correlation and copulas are unable to capture cause-effect relationships among random variables, and they also cannot make use of available knowledge about causal structures [11]. This is why the process of capturing conventional stochastic dependencies using copula functions is referred to as "reduced-form, black-box models, which do not provide insight into the directional dependencies or fundamental risk drivers" that govern industrial processes [11].

Another ongoing challenge of the application of copula functions is the difficulty in their application for high-dimensional problems. The research contributions to overcome this challenge are either limited to a mixture of trees compositions [23], or rely on a recursive construction of conditional bivariate copulas (also known as vine copulas) [24,25]. However, these approaches are elaborate for high dimensions and their applications are limited to a relatively small number of variables [5].

Table 8.1 summarizes recent attempts to develop multivariate copula-based graphical modelling approaches. Among these models, the CBN model proposed by Elidan [5] is selected in this work as it provides an innovative tool to decompose distributions associated with a DAG because: (i) it is flexible enough to use any higher variate copulas; (ii) it provides control of marginal distributions; (iii) its practical applications are relatively simple as it uses the same graphical structure as BN. The CBN model uses a copula decomposition of distributions associated with a DAG. In CBN analysis, the local copula density function together with marginal distributions can be used to parameterize a

conditional density required to build BNs. This offers great flexibility in modeling high-dimensional continuous distribution while retaining copula advantages. The following section provides a methodology based on the CBN model for multivariate probabilistic safety analysis of process facilities.

Table 8.1. Multivariate copula-based graphical modelling methods; adopted from Elidan [26]

Model/References	Structure	Copula	Variables
Vines [24,25]	Conditional dependence	Any bivariate	< 10 in practice
Nonparametric BBN [27–29]	BN plus vines	Gaussian in practice	100s
Tree-averaged [23,30]	Mixture of trees	Gaussian	10s
Copula networks [5]	BN	Any	100s

8.3. Methodology

As shown in Figure 8.2, the proposed methodology for probabilistic analysis of abnormal operational conditions using CBN models consists of two main steps. Step 1 is an off-line process that identifies the network nodes and the CBN model structure. Step 2 is an on-line process of inference analysis that uses the developed CBN model to estimate the real-time probability of abnormal conditions using new evidence from the system. The details of each step are discussed in the following sections.

As an example application, the estimated real-time probability values can be used to analyze warnings and conduct root-cause diagnosis of abnormal process conditions. These latter steps are shown with dashed lines in Figure 8.2 as they are not the main focus of this paper.

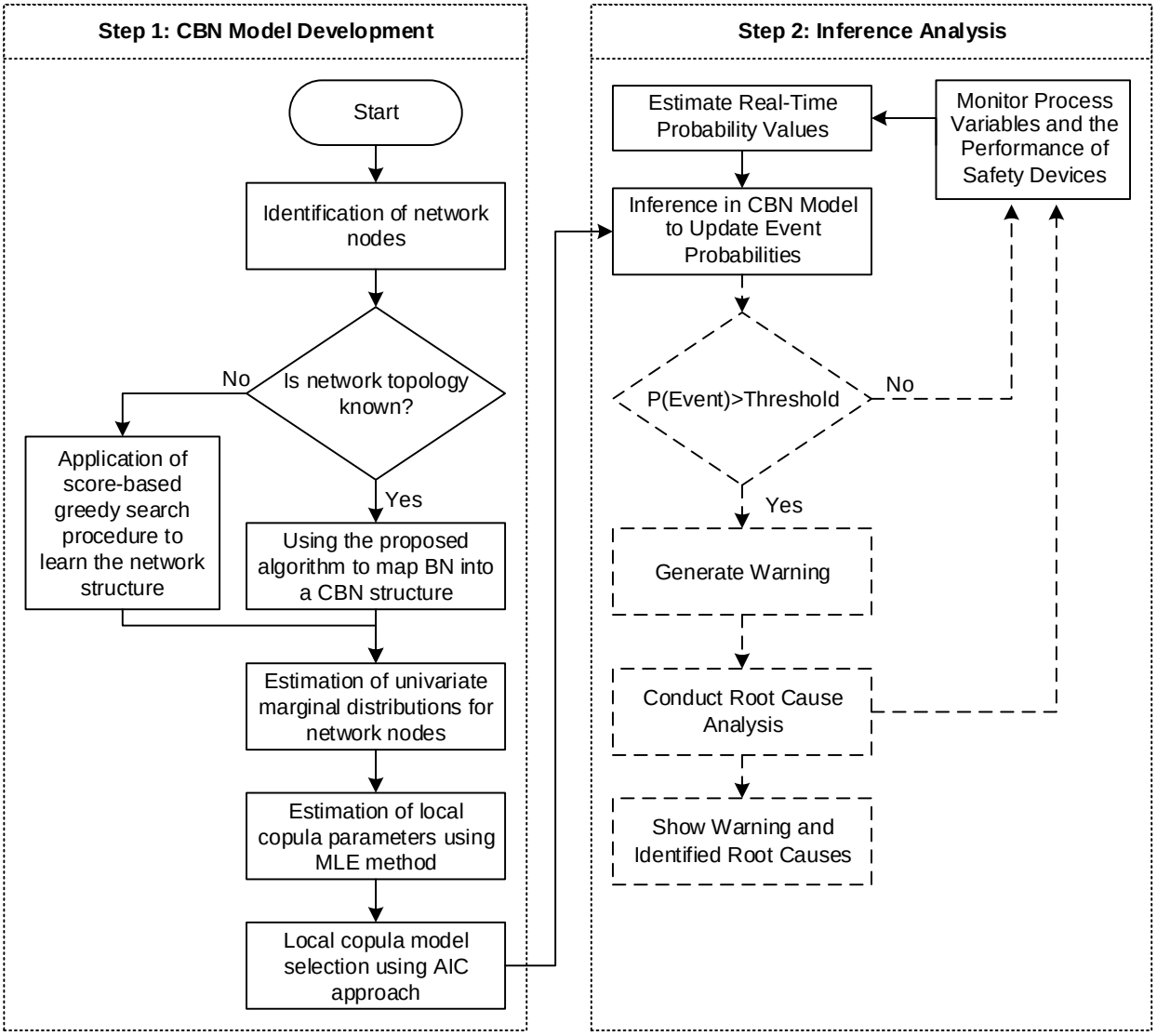


Figure 8.2. Proposed methodology for development and application of the CBN model in safety analysis

8.3.1. Step 1: CBN Model Development

8.3.1.1. Step 1.1: Identification of Network Nodes

The first step to develop the CBN model is to identify the network nodes, which represent system variables in the network. Table 8.2 defines four types of nodes that are used in construction of the CBN structure, along with a few examples.

Table 8.2. Node types for construction of a CBN; adopted from Ahmed et al. [31]

Node Type	Associated Process Attributes	Description	Examples
Root	Root causes	Root causes and/or process faults that influence the process deviations.	External disturbances, equipment malfunctions, control system failures or human errors.
Intermediate Type I	Symptoms	Deviation of process characteristics from their target values.	Deviation of temperature and differential pressure of a distillation column from their operating limits.
Intermediate Type II	Scenarios	Process operating conditions that influence an event.	Failure of the control system and operator to detect and correct process symptoms.
Leaf	Events	Undesirable abnormal process conditions.	Product quality degradation, distillation column flooding, and reactor runaway.

The Hazard and Operability Study (HAZOP) can be used to identify the potentially significant events. Then, the maximum credible accident scenario identification [32] may be used to envisage the potential scenarios leading to each identified abnormal event. Subsequently, the information from the HAZOP study is used to identify root causes of each event. Alternatively, Failure Mode and Effect Analysis (FMEA) can be used to identify possible root causes. Finally, the process variables, the deviation of which can result in the identified events, are selected [31]. Detailed description of these methods is not within the scope of this work.

8.3.1.2. Step 1.2: Network Topology Development

Like BN analysis, a CBN model takes advantage of a graphical structure to represent the causality among random variables. Additionally, CBN analysis uses copula functions to capture all of the dependence information among random variables. In general, the topology of a CBN structure remains the same as the topology of the corresponding BN

structure. Thus, the construction of the network structure depends on the causal relationships among variable nodes [7]. However, rather than the conditional probability tables in a BN model, local copulas are used in a CBN model to capture the dependence structure and strength of the relationship among nodes. Therefore, if the BN topology of a given system is known, the same graph structure can also be used for the corresponding CBN model.

The mapping algorithm shown in Figure 8.3 can be used to map the existing BN model into a CBN model. As shown in Figure 8.3, the nodes structure remains the same in both BN and CBN. However, in contrast to the BN, the equivalent CBN model will have flexibility in terms of assigning marginal distributions to each node. Finally, local copulas are used to represent dependence structure among variables.

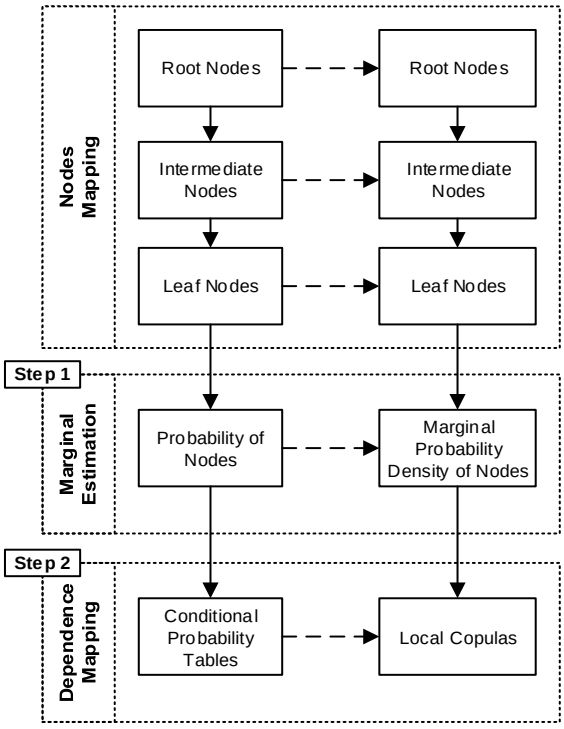


Figure 8.3. Mapping BN to CBN

If the BN structure of a given system is unknown, a search process can be used to find the most useful network structure which can represent the given dataset. Different search algorithms have been proposed to rank the candidate network structures based on an estimated score, among which the greedy search algorithm is a common choice. Eban & Elidan [33] proposed a standard greedy search algorithm that can be used to apply local structure modifications (e.g., add/delete/reverse edge) based on a model selection score. They suggested using the Bayesian Information Criterion (BIC) that, like other scores, balances the likelihood and the complexity of the model. For complex networks with a high number of variables, however, calculating the score for each of the numerous candidate structures is computationally demanding. Elidan [34] and Tenzer & Elidan [35] showed that the expected likelihood of an edge in the model is monotonic in the magnitude of Spearman's correlation coefficient, dented by ρ_S, for two important copula families. They also showed numerically that this relationship holds for many other popular copulas. Motivated by this result, the empirical ρ_s can be used as a model selection measure to crudely yet efficiently pre-rank candidate structure modifications. Then, more precise, but costly, computation of the BIC score can be performed for only the most promising candidates [34].

This work assumes that the causal relationship among network variables for a given system is already known, to keep the work focused on CBN model development. An interested reader may refer to Tenzer & Elidan [35] to learn more about the score-based network structure search procedure. Application of the proposed CBN methodology along with the search procedure described above ensures assessment of potential hidden correlations (as

measured via Spearman's ρ_s) and causal relationships (as determined by identifying the potential parents) among variables.

8.3.1.3. Step 1.3: CBN Model Development

The CBN model proposed by Elidan [5] is used in this work to decompose a joint distribution associated with a DAG. Let G be a DAG with nodes corresponding to the set of random variables $\boldsymbol{X} = \{X_1,\ldots, X_d\}$, and let $\mathbf{pa}_i = \left\{\mathbf{pa}_{i1},...,\mathbf{pa}_{ik_i}\right\}$ be the parents of X_i in G. Elidan [5] defined the CBN model as a triplet $\mathcal{D} = (G,\Theta_C,\Theta_f)$ that represents $f_X\left(\boldsymbol{x}\right)$, with lower case letters denoting assignment to variables. G encodes the independencies ($X_i \perp Nd_i \mid \mathbf{pa}_i$), which are assumed to hold in $f_X\left(\boldsymbol{x}\right)$, where $\perp$ denotes the independence relationship, and Nd_i are nodes that are not descendants of X_i in G. Θ_C is a set of local copula functions $C_i\left(F\left(x_i\right),F\left(\mathbf{pa}_{i1}\right),...,F\left(\mathbf{pa}_{ik_i}\right)\right)$ that is associated with the nodes of G that have at least one parent. In addition, Θ_f is the set of parameters representing the marginal densities $f_i(x_i)$ (and distributions $F_i(x_i)$). For compactness, in this work $f\left(x_i\right) \equiv f_{X_i}\left(x_i\right)$ and $F\left(x_i\right) \equiv F_{X_i}\left(x_i\right)$. Then, the joint density $f\left(\boldsymbol{x}\right)$ can be shown as:

$$f\left(\mathbf{x}\right) = \prod_i R_{c_i}\left(F\left(x_i\right),\left\{F\left(\mathbf{pa}_{ik}\right)\right\}\right) f\left(x_i\right) \tag{8.2}$$

where, if X_i has at least one parent in the graph G, the term $R_{c_i}\left(F\left(x_i\right),\left\{F\left(\mathbf{pa}_{ik}\right)\right\}\right)$ denotes the conditional copula density and is defined as:

$$R_{c_i}(\cdot) = \frac{c_i\left(F(x_i), F(\mathbf{pa}_{i1}), ..., F(\mathbf{pa}_{iK_i})\right)}{\frac{\partial^K C_i\left(1, F(\mathbf{pa}_{i1}), ..., F(\mathbf{pa}_{iK_i})\right)}{\partial F(\mathbf{pa}_{i1}) ... \partial F(\mathbf{pa}_{iK_i})}}. \quad (8.3)$$

When X_i has no parents in G, $R_{c_i}(\cdot) \equiv 1$. The term $R_{c_i}\left(F(x_i), \{F(\mathbf{pa}_{ik})\}\right)$ is always a valid conditional density, namely $f(x_i|\mathbf{pa}_i)$, and can be easily computed. In particular, when the copula density $c(\cdot)$ has an explicit form, so does $R_{c_i}(\cdot)$, since it involves derivatives of a lesser order [5].

Like the BN framework, the foundation of a CBN model is a local conditional density. However, in a CBN model the conditional densities are parametrized using copulas according to the following lemma:

Lemma 1: Let $f(x|\boldsymbol{y})$, with $y = \{y_1,\ldots,y_K\}$, be a conditional density function and let $f(x)$ be the marginal density of X. Then there exists a copula density function $c(F(x), F(y_1), ..., F(y_K))$ such that:

$$f(x|\boldsymbol{y}) = R_c\left(F(x), F(y_1), ..., F(y_K)\right) f(x). \quad (8.4)$$

Thus, any copula density function $c(x, y_1, \ldots, y_K)$, together with $f(x)$, can be used to parameterize a conditional density $f(x|\boldsymbol{y})$. See Elidan [5] for the proof.

8.3.1.4. Step 1.4: Copula Selection and Parameter Learning

The decomposable form of the joint density defined by the CBN model in Equation (8.2) facilitates relatively efficient copula estimation using standard approaches such as

maximum likelihood evaluation. However, from an estimation perspective, performing the decomposable estimation can be complicated, as the univariate marginals in a CBN model are usually shared across the entire model. The frequently used technique in the copula community to overcome this challenge is to first estimate the marginals and then learn the copula parameters [33]. Therefore, learning the structure of a CBN model involves three steps:

(i) identification of marginal distributions;

(ii) estimation of copula parameters; and

(iii) selection of the best fitting copula.

Appendix 8.A provides details of implementing this three-step learning process. Methods based on maximum likelihood evaluation and Akaike's Information Criterion (AIC) are presented in Appendix 8.A to estimate copula parameters and rank the competing candidate copulas for a given CBN structure.

8.3.2. Step 2: Inference Analysis

The developed CBN model can be used to perform probabilistic inference for updating the prior occurrence probability of events given new information by adopting the BN inference analysis. Let $\boldsymbol{E}$, $\boldsymbol{S}$ and $\boldsymbol{A}$ be a finite set of real-valued random variables denoting events, symptoms (evidence) and root cause nodes of a network, respectively. Using copula parameterization of the conditional density in Lemma 1, Equation (8.4), the CBN inference analysis can then be shown as:

$$f\left(E_j|\boldsymbol{S}\right)=R_c\left(F\left(E_j\right),F\left(S_1\right),...,F\left(S_K\right)\right)f\left(E_j\right) \tag{8.5}$$

where $f\left(E_j|\boldsymbol{S}\right)$ denotes the posterior probability of event E_j given the observation of certain symptoms; $f\left(E_j\right)$ is the prior distribution of event E_j; j counts the number of events; and $R_c\left(\cdot\right)$ denotes the conditional copula density and is defined in Equation (8.3). Hence, the updated and/or real-time probabilities of symptom nodes can be plugged into Equation (8.5) to obtain the updated event probabilities. The real-time probabilities of symptom nodes can be estimated based on the type of symptom. For instance, Bao et al. [36] proposed the application of a three-sigma rule to evaluate the deviation probability of monitored process variables using real-time process measurements. As an another example, Abimbola et al. [37] used physical reliability models of constant strength and exponentially distributed random stress to estimate the real-time failure probability of drilling equipment as a function of drilling depth.

The updated event probability can be continuously compared to a threshold probability. Therefore, once the probability of the event occurrence exceeds the threshold, the event warning is annunciated to inform the operator about the unsafe condition. For cases in which the probability of the event occurrence exceeds the threshold, the copula parameterization of the conditional density using the developed CBN model can be used to conduct backward analysis to update the probability of root-cause nodes:

$$f\left(A_r|\boldsymbol{S}\right)=R_c\left(F\left(A_r\right),F\left(S_1\right),...,F\left(S_K\right)\right)f\left(A_r\right) \tag{8.6}$$

where A denotes root-causes, r counts the number of root-causes and S denotes symptoms (evidence). $f\left(A_r|\boldsymbol{S}\right)$ indicates the occurrence probability of a particular root cause given the observation of certain evidence. The estimated probability values using Equation (8.6)

can be used to rank the contribution of identified root causes in the occurrence of the observed event.

8.4. Case study: Managed Pressure Drilling

8.4.1. Case Study Description

To demonstrate the application of the proposed CBN model in complex real-life problems, both BN and CBN models are applied to an offshore managed pressure drilling (MPD) operation case study, adopted from a study by Abimbola et al. [38]. MPD is an adaptive drilling process used to precisely control the annular pressure profile throughout the wellbore. The overbalanced drilling technique used in the MPD avoids the flow of formation fluid into the wellbore [38]. In their study, Abimbola et al. [38] proposed a methodology based on the BN approach for safety and risk analysis of the MPD operation. The developed BN is analyzed to assess the safety critical elements of constant bottom-hole pressure drilling techniques and their safe operating pressure regime.

Figure 8.4 shows the event tree for the underbalanced drilling scenario. Insufficient mud weight, unexpected pore pressure and lost circulation, in conjunction with a failure of the MPD system in preventing the underbalance, are among the main causes of an underbalanced drilling scenario [38]. As shown in Figure 8.4, the unexpectedly high pore pressure (P_P) beyond the bottom hole pressure (P_{BH}) is considered as the initiating event for an underbalanced scenario. The event tree in Figure 8.4 also demonstrates four safety barriers which have been considered to prevent the consequences of the underbalanced

scenario. For simplicity, analyzing the fault tree associated with the causes of initiating an underbalanced drilling scenario is not included in the scope of this case study.

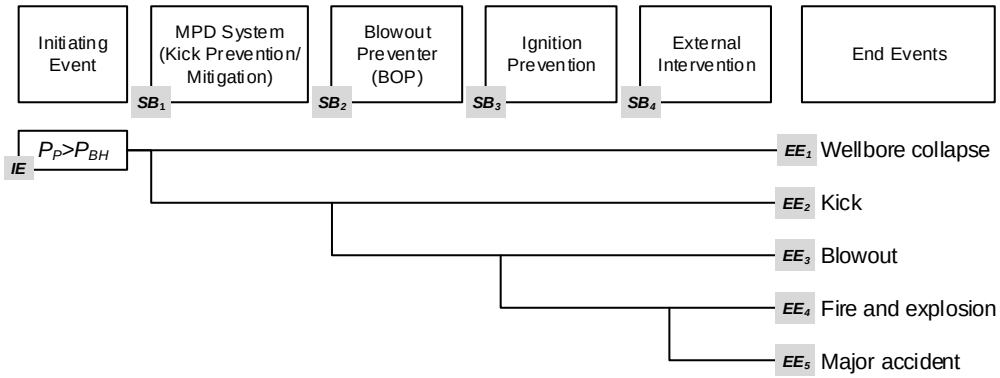


Figure 8.4. Event tree and safety barriers for underbalanced drilling scenario; adopted from Abimbola et al. [38]

In this case study, first a BN is constructed for the underbalanced drilling scenario. Then, the developed BN is mapped into a CBN model. The developed CBN is then used to conduct inference analysis to update the probability of end events as new evidence from the drilling operation becomes available. Finally, the appealing features of the proposed CBN methodology to address the inherent shortcomings of the BN approach are discussed.

8.4.2. BN Model Development

Figure 8.5 shows separate BNs developed for each state of the consequence node of the event tree of the underbalanced drilling scenario (in Figure 8.4). The descriptions of different nodes in Figure 8.5 are provided in Tables 8.3 and 8.4. The developed BNs represent the relationships between potential consequences and safety barriers, the failure of which affect the probability of each consequence state. Considering each consequence

state (event) as a separate node allows consideration of each state as a continuous variable while developing the CBN model in the next section.

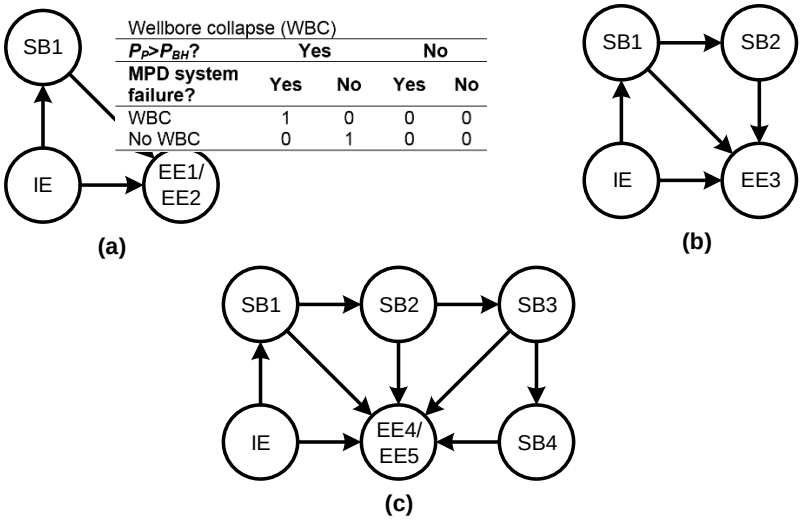


Figure 8.5. The BN models for the underbalanced drilling scenario; (a) wellbore collapse (EE1) and kick (EE2), (b) blowout (EE3), (c) fire/explosion (EE4) and major accident (EE5)

As an example, the consequence state EE_5 in Figure 8.5(c), denoting catastrophic fire and explosion, is selected to analyze the BN model. Given a family $\boldsymbol{V} = (IE, SB_i, EE_5)$, i = 1,2,3,4, of variables associated to a DAG G_{BN} in Figure 8.5(c), the joint distribution of $\boldsymbol{V}$ is developed using the BN model. Then, conditional probability tables (CPT) are assigned to the consequence nodes as well as to the safety barrier nodes. For instance, the CPT for the Wellbore Collapse (EE_2) end event is embedded in Figure 8.5(a). The CPT for consequence state nodes acts like a logical AND-gate where values 1 and 0 represent occurrence or non-occurrence of the associated event in the CPT. The CPT of the safety barrier nodes similarly takes 0 and 1 values to represent failure and success of each safety barrier node.

The probability values in Table 8.3, sourced from Abimbola et al. [38], and the developed CPT tables are plugged into the BN model using the GeNIe modeling environment

developed by the Decision Systems Laboratory of the University of Pittsburgh and available at http://genie.sis.pitt.edu/. The calculated values of probability of occurrence of each consequence state are shown in Table 8.4, which match the results in Abimbola et al. [38].

Table 8.3. Probability values for the initiating event and safety barriers failure on demand

Node	Description	Probability
IE	Underbalanced condition (the initiating event)	9.75E-03
SB1	MPD system failure	8.14E-02
SB2	Blowout preventer (BOP) failure	7.00E-04
SB3	Ignition prevention failure	1.07E-01
SB4	External intervention (fire-fighting, evacuation, drilling of relief well, etc.) failure	2.71E-02

Table 8.4. Underbalanced scenario predictive frequency of occurrence

End event	Description	Estimated probability
EE0	Near balanced condition	9.90E-01
EE1	Wellbore collapse	8.96E-03
EE2	Kick	7.93E-04
EE3	Blowout	4.96E-07
EE4	Explosions, fire, major injury to some fatalities, minimal environmental pollution	5.78E-08
EE5	Catastrophe (fatalities, loss of rig, major environmental damage)	1.61E-09

8.4.3. CBN Model Development

8.4.3.1. Problem Formulation

In this section, the MPD case study is expanded into a more general case to highlight the strengths of the proposed CBN model compared to BN analysis. For this purpose, the probability of each network node is considered to follow the distributions in Table 8.5, rather than the deterministic point-based probability values used in Abimbola et al. [38]. The distribution parameters in Table 8.5 are selected in such a way that the mean of each

distribution equals the probability value of the associated node reported in Tables 8.3 and 8.4. The probability distributions are selected from different families to show the flexibility of the CBN model. Exact estimation of marginal failure probability distributions is not within the scope of this work. For specific applications, methodologies discussed in Appendix 8.A.1 can be applied to estimate univariate marginals.

It should be noted that in the CBN model, the continuous random numbers represent the probabilities of the failure, not the probabilities of the events. Therefore, different notation has been used in Table 8.5 to show each node where, for example, p_{SB1} shows the probability of failure of SB_1. For simplicity, it is assumed that the causal structure of the original random variables IE, SB_1, ..., SB_4 and EE_5 still holds for p_{IE}, p_{SB1}, …, p_{SB4} and p_{EE5}. The validity of this assumption can be investigated using causality analysis techniques described in [7].

Table 8.5. Probability distributions for the MPD case study

Node	Description	Probability distributions
p_{IE}	Underbalanced condition	Normal; μ_N = 9.75E-03*, σ_N = 0.01*
p_{SB1}	MPD system	Weibull; β_w = 1*, θ_w = 8.14E-02*
p_{SB2}	BOP	Weibull; β_w = 1, θ_w = 7.00E-04
p_{SB3}	Ignition prevention	Normal; μ_N = 1.07E-02, σ_N = 0.04
p_{SB4}	External intervention	Lognormal; μ_{LN} = log(2.71E-02)*, σ_{LN} = 0.1*
p_{EE1}	Wellbore collapse	Gamma; k_g = 1*; α_g = 8.96E-03*
p_{EE2}	Kick	Gamma; k_g = 1; α_g = 7.93E-04
p_{EE3}	Blowout	Gamma; k_g = 1; α_g = 4.96E-07
p_{EE4}	Fire and explosion	Gamma; k_g = 1; α_g = 5.78E-08
p_{EE5}	Catastrophe	Gamma; k_g = 1; α_g = 1.61E-09

* μ_N and σ_N are mean and standard deviation of Normal distribution; μ_{LN} and σ_{LN} are mean and standard deviation of Lognormal distribution; β_w and θ_w are shape and scale parameters of Weibull distribution; k_g and α_g are shape and scale parameters of Gamma distribution.

8.4.3.2. Training Dataset

Having estimated the probability distributions, a training dataset is generated with a presumable dependence structure to facilitate the comparison of BN and CBN models. In real-life applications, data from process history and safety system performance can be used as the training dataset. However, to emphasize the generic power of the CBN model, a simulated training dataset is used in this case study. For simplicity, the entire training dataset is generated using the *t*-copula by following these two steps:

i. Simulate a realization of multivariate random vector $\bar{u}$ with marginals uniformly distributed on [0 1] from a *t*-copula. This simulation is conducted by construction of a multivariate *t* distribution, followed by its transformation using the corresponding *t* cumulative distribution function (CDF). The correlation coefficient values and the degree of freedom required to parametrize the *t*-copula are discussed below.

ii. Transform back the simulated copula random numbers using the corresponding inverse of the cumulative distribution function of each node in Table 8.5.

This two-step transformation creates dependent random numbers representing the network nodes with a presumable dependence structure that can be used to test the presented copula learning methodology. A sensitivity analysis is provided in Section 8.4.5 to investigate the effect of a changing copula family or copula parameter as well as the effect of noise on estimated probability values.

Table 8.6 shows the correlation coefficient values which are used to parametrize each local *t*-copula. 5 degrees of freedom are considered for all *t*-copulas to allow for heavy-tailed

distributions. To facilitate visual representation of the dependence structure among nodes, Figure 8.6, that has scatter plots to demonstrate joint distribution of different node pairs, is also included. Only p_{EE5} is included in Figure 8.6 to simplify the figure. The plots on the diagonal of Figure 8.6 show the marginal distributions of each node and the lower-left panel shows the correlation coefficient values, taken from Table 8.6. The generated dataset is used in the next section to demonstrate the application of the presented copula learning methodology.

Table 8.6. Upper-right panel: Correlation coefficient values used to generate the training dataset for the MPD case study. Lower-left panel: Estimated *t*-copula parameters (estimated values are shown in italic format)

Nodes	Parents	Nodes									
		p_{IE}	p_{SB}	p_{SB2}	p_{SB3}	p_{SB4}	p_{EE1}	p_{EE2}	p_{EE3}	p_{EE4}	p_{EE5}
p_{IE}	Not applicable (NA)	1	0.75	0.28	0.30	0.15	0.68	0.49	0.50	0.46	0.52
p_{SB1}	p_{IE}	*0.750*	1	0.53	0.50	0.21	0.55	0.50	0.52	0.49	0.48
p_{SB2}	p_{SB1}	*0.275*	*0.524*	1	0.63	0.20	0.52	0.52	0.54	0.48	0.52
p_{SB3}	p_{SB2}	*0.299*	*0.496*	*0.627*	1	0.47	0.50	0.60	0.58	0.50	0.75
p_{SB4}	p_{SB3}	*0.144*	*0.205*	*0.194*	*0.469*	1	0.50	0.61	0.60	0.65	0.78
p_{EE1}	p_{IE}, p_{SB1}	*0.678*	*0.546*	*0.517*	*0.499*	*0.494*	1	0.62	0.50	0.50	0.65
p_{EE2}	p_{IE}, p_{SB1}	*0.490*	*0.497*	*0.515*	*0.599*	*0.608*	*0.617*	1	0.78	0.70	0.65
p_{EE3}	p_{IE}, p_{SB1},p_{SB2}	*0.499*	*0.517*	*0.537*	*0.581*	*0.599*	*0.500*	*0.780*	1	0.86	0.78
p_{EE4}	p_{IE}, p_{SB1},…, p_{SB4}	*0.458*	*0.486*	*0.474*	*0.499*	*0.651*	*0.498*	*0.700*	*0.860*	1	0.80
p_{EE5}	p_{IE}, p_{SB1},…, p_{SB4}	*0.517*	*0.476*	*0.517*	*0.752*	*0.778*	*0.647*	*0.651*	*0.780*	*0.800*	1

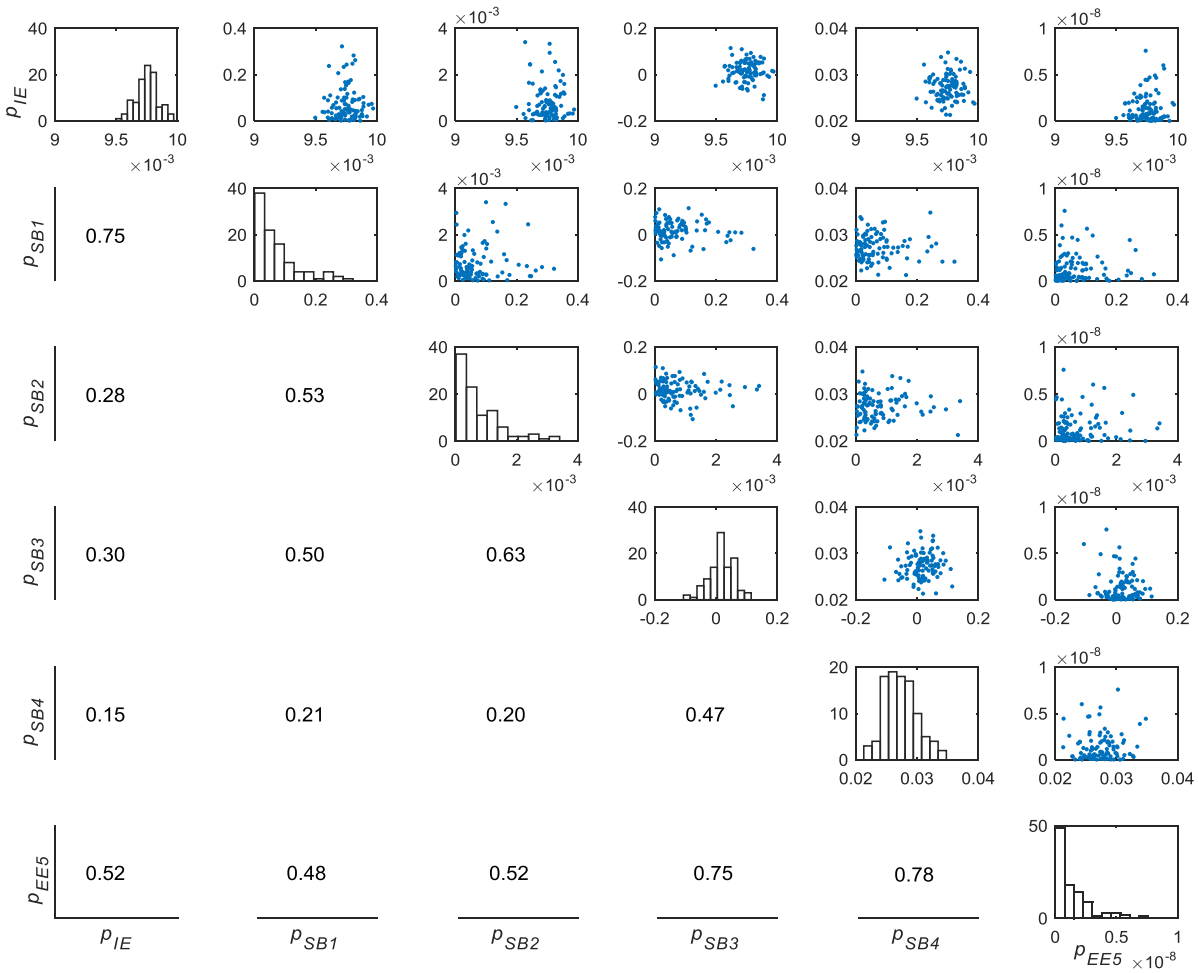


Figure 8.6. A part of the training dataset including marginal distributions (diagonal), joint distributions (upper right panel), and correlation coefficient values (lower left panel) for the MPD case study

8.4.3.3. Copula Learning

The copula learning methodologies presented in Appendix 8.A are used to select local copulas and associated parameters to capture the dependence structure among each network node and its parents. Table 8.6 shows the parents for each node along with the correlation coefficient values for each local *t*-copula (shown in the lower left of Table 8.6 in italic format), estimated using the ML method with a sample size of 2,000. Comparing the original and the estimated ρ values in Table 8.6, one can see the efficiency and high

accuracy of the ML method in estimating the copula parameters, even for a relatively small sample size of 2,000.

The next step is to select the local copulas. In principle, the CBN allows application of any family of local copulas in the model and even mixes different copula families without significant computational difficulty. The AIC approach is used to select each local copula and the results are shown in Table 8.7 for a sample size of 2,000. Other than the *t*-copula, the Clayton and Gumbel copulas from the Archimedean family are also selected as candidate copulas as they are able to represent dependency in distribution tails.

From the calculated Akaike weights (shown in Table 8.7) using Equation (8.A5) in Appendix 8.A, it can be seen that, even for a relatively small sample size of 2,000, the AIC approach has been able to select the *t*-copula as the best approximating model since its Akaike weight is significantly greater than the others. The Akaike differences (Δ_i) can be used as a criterion to compare the level of empirical support for each model. However, when using the AIC approach, the candidate models with a value of $\Delta_i > 10$ should not be considered as competing models [39]. The large value of the estimated Δ_i for Clayton and Gumbel copulas shows that, for this case study, they cannot even be considered as competing models compared to *t*-copula.

Table 8.7. Parameter estimations and Akaike weights for the MPD case study

Node	Parent	Copula	$\hat{\delta}_i$	$\ell(\delta_i)$	AIC$_i$	Δ_i	w_i
p_{SB1}	p_{IE}	*t*	0.74	431.50	-859.00	0.00	1.00
		Clayton	1.60	315.68	-629.37	229.63	0.00
		Gumbel	2.06	401.35	-800.69	58.30	0.00
p_{SB2}	p_{SB1}	*t*	0.55	182.74	-361.49	0.00	1.00
		Clayton	0.87	129.73	-257.46	104.02	0.00
		Gumbel	1.54	168.25	-334.51	26.98	0.00
p_{SB3}	p_{SB2}	*t*	0.64	302.71	-601.42	0.00	1.00
		Clayton	1.10	219.72	-437.44	163.98	0.00
		Gumbel	1.70	279.15	-556.31	45.11	0.00
p_{SB4}	p_{SB3}	*t*	0.46	137.86	-271.73	0.00	1.00
		Clayton	0.66	90.87	-179.73	91.99	0.00
		Gumbel	1.41	121.01	-240.02	31.70	0.00

8.4.4. Inference Analysis

As an example, Figure 8.7 shows the graphical structure of the developed CBN model for catastrophe conditions (EE_5) in the MPD case study, where C_i represents local copulas. The developed CBN helps to clearly establish both causality and correlation among network nodes, an important advancement compared to traditional graphical modelling approaches. In Figure 8.7, similar to the equivalent BN in Figure 8.5(c), the cause and effect relationships between network nodes are identified using connecting arrows. Moreover, the local copulas and related parameters in Figure 8.7 help to describe the structure, size and direction of dependency among network nodes.

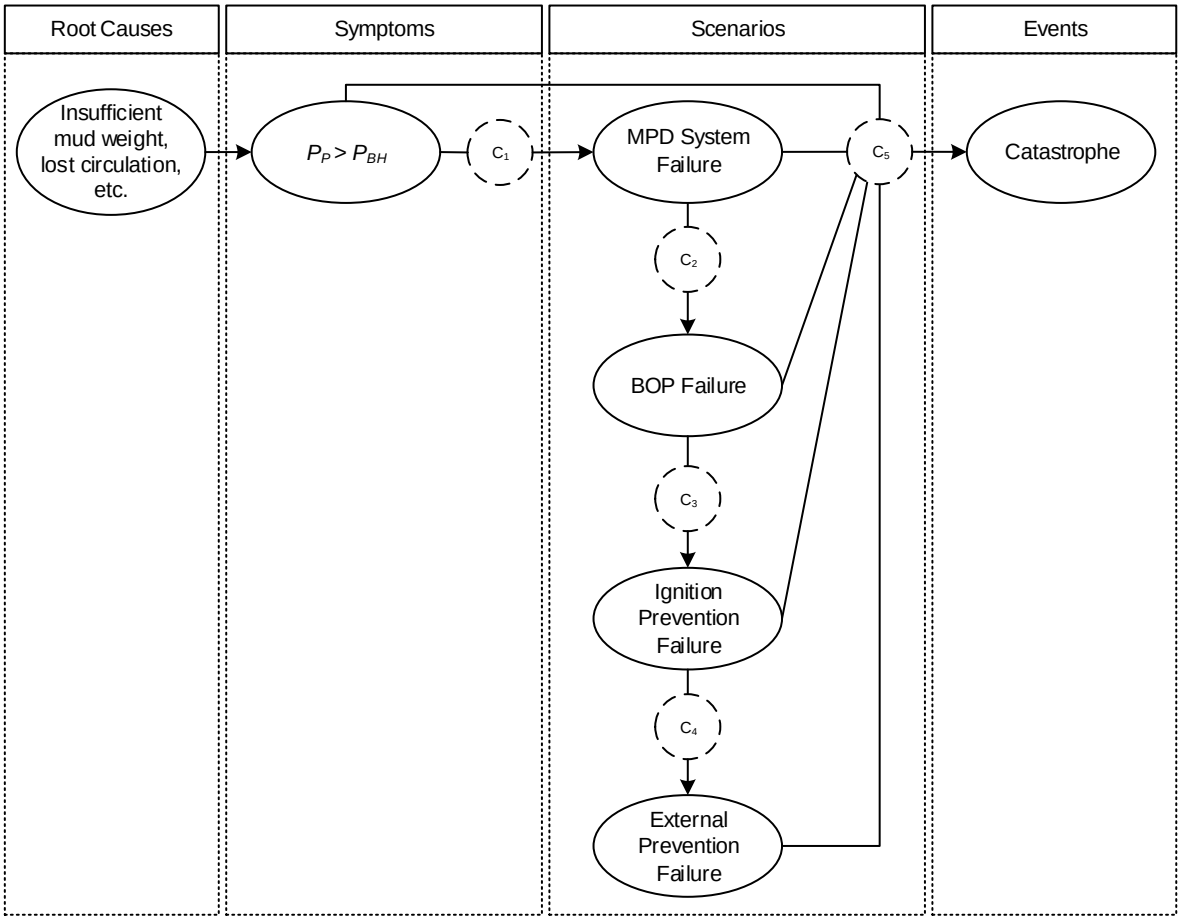


Figure 8.7. The CBN model for the catastrophe conditions (EE5) in the MPD case study

The developed CBN model is used to conduct inference analysis and revise the probability of different end states (p_{EEi}) as the probability of underbalanced conditions (p_{IE}) increases as a function of the drilling depth. Abimbola et al. [37] used the following physical reliability model of constant strength and exponentially distributed random stress to estimate the probability of failure of drilling equipment (p_{DE}):

$$p_{DE} = \exp\left(-S_\sigma / \left(E\left(P_P\right) - 0.052 \times ECD \times h\right)\right) \tag{8.7}$$

where S_σ is the rated strength of the drilling equipment in pounds per square inch (*psi*); $E(P_P)$ is the expected value of the measured pore pressure (stress) in *psi*; *ECD* is the Equivalent Circulating Density in pounds per gallon (*ppg*) comprising the mud hydrostatic pressure and the frictional pressure loss in the annulus; and *h* is the drilling vertical height in *ft*. Considering a salt water formation fluid for this case study, $E(P_P)$ can also be expressed as a function of *h* as $E(P_P) = 0.465h$ [37]. Assuming that the failure of the drilling equipment causes the underbalanced conditions, Equation (8.7) is used in this work to estimate the real-time probability of the underbalanced conditions (p_{IE}) as a function of *h*. Then, the proposed inference analysis methodology in Equation (8.5) is used to revise the probability of different end states (p_{EEi}) using the developed CBN model.

To demonstrate the application of the proposed inference analysis methodology using the CBN model, a case is considered where *h* = 16,900 *ft* and the drilling rate is 50 *ft* per hour. To account for the potential measurement variability, the pore pressure, stress, is considered to follow a Normal distribution with the mean value of $\mu_{P_p} = 0.465h$ and standard deviation of $\sigma_{P_p} = 108.25$ *psi*. A gasified drilling fluid of density 3.5 *ppg* is considered in this case study. Figure 8.8 shows the P_P time-plot of the drilling operation. Up to 11:32 AM, the P_P increases steadily due to the increase in *h*. At 11:32 AM, it is assumed that the faulty conditions in the drilling equipment resulted in a 20% increase in σ_{P_p}. Other than the increase in P_P as a function of *h*, an unexpected 100 *psi* increase in P_P also occurs from 11:32 AM to 11:55 AM due to the faulty conditions, moving the operation closer to underbalanced drilling conditions. Accordingly, the probability of underbalanced conditions increases as a function of *h*, which can be estimated using Equation (8.7). At

11:55 AM, the second faulty conditions occurred, which resulted in an additional 20% increase in σ_{P_P} and an unexpected 200 *psi* increase in P_P from 11:55 AM to 12:08 PM.

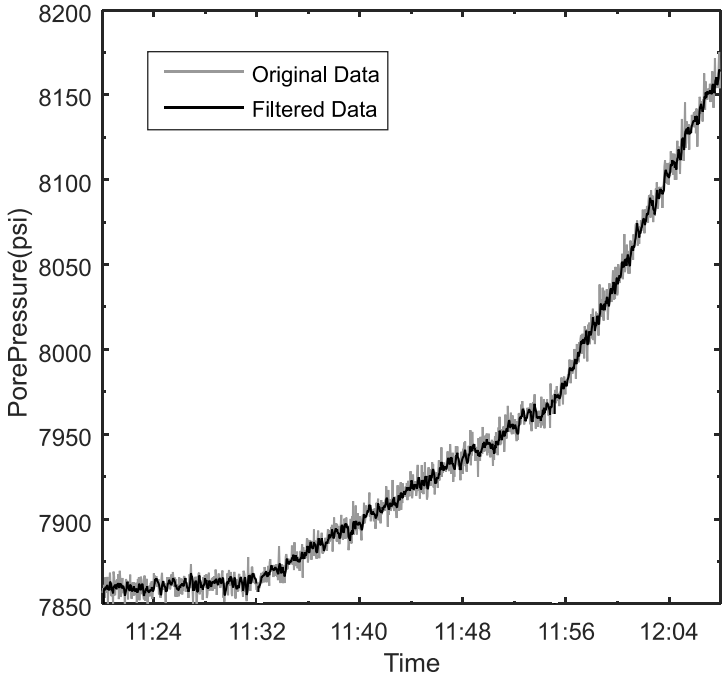


Figure 8.8. Pore pressure time-plot

Assuming that the P_P measurements are intended to be used for warning generation, noise filtering is carried out for pressure measurements using the moving average filter in order to minimize false warnings [40]. The filtered measurements are shown in darker colour in Figure 8.8. The estimated probabilities of underbalanced conditions using Equation (8.7) are plugged into the developed CBN model as the evidence to conduct inference analysis using Equation (8.8). As an example, Figure 8.9 shows the estimated posterior probability of a blowout event given the change in P_P over h and the consequent increase in the probability of underbalanced conditions.

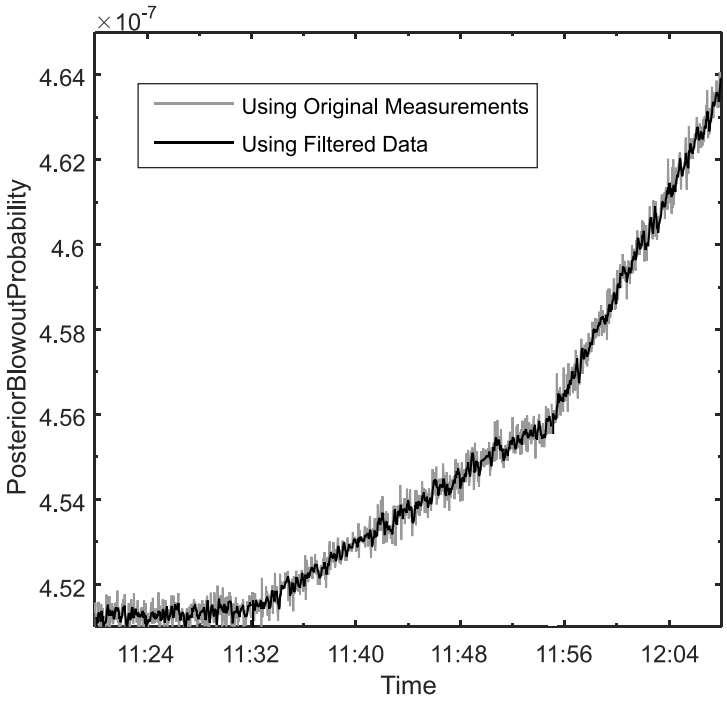


Figure 8.9. Posterior blowout probability time-plot

Similarly, Equation (8.9) can be used to conduct root cause analysis by revising the probability of identified root causes. For conciseness, the root cause analysis is excluded from the scope of this case study. As shown in this case study, using the proposed CBN methodology, noisy process measurements with any marginal distributions and complex non-linear dependence structure can be plugged into the model to revise the probability of the events on a real-time basis. By assigning a threshold value to the probabilities of different end events, the estimated probability time-plot in Figure 8.9 can be used for alarm analysis and annunciation. This provides a flexible framework for the event-based early warning system approach by assigning warnings to undesired events rather than assigning alarms to each individual monitored variable, resulting in reduction of the probability of alarm flooding. Further work on application of the proposed model for warning generation

and fault diagnosis and its integration with loss modelling to develop a risk-based event-based early warning system are the subjects of ongoing research by the authors.

8.4.5. Sensitivity Analysis

To account for the cases when the copula selection approach chooses a wrong copula for a given dataset, an experiment is performed using *t*, Clayton and Gumbel copulas to estimate the posterior probability of blowout (EE_2) to regenerate Figure 8.9 with different copulas. The correlation coefficient value of $\rho = 0.49$ between p_{IE} and p_{EE2}, selected from Table 8.6, is used to parametrize the *t*-copula. To parametrize the Clayton and Gumbel copulas, Kendall's τ, estimated as $\tau = (2/\pi)\arcsin\rho$ [6], is used. As can be seen from Figure 8.10, the posterior blowout probabilities estimated from three different copulas are significantly different. In Figure 8.10, the posterior blowout probability time-plot estimated using *t*-copula matches Figure 8.9. However, application of Clayton and Gumbel copulas have resulted in relatively significant overestimation and underestimation of the probability values, respectively, as change in the type of copula causes a potentially significant change in the dependence structure among network nodes. Although this may be concluded to be a shortcoming of the CBN approach, the sensitivity analysis investigations conducted in the rest of this section show that this is not a concern when using the proposed AIC copula selection methodology.

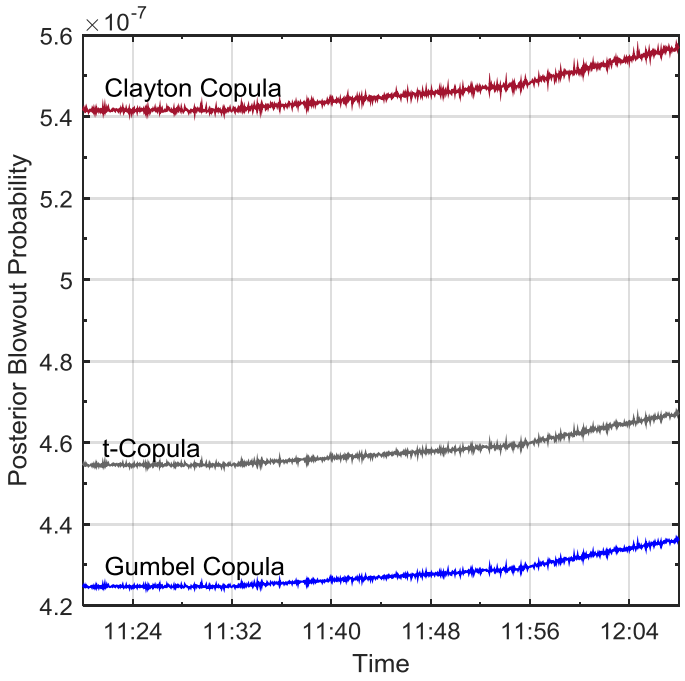


Figure 8.10. Posterior blowout probability time-plots using three copulas

A sensitivity analysis is performed to investigate the performance of the proposed AIC copula selection methodology. Random noise from different sources for a given dataset can hinder identification of the true dependence structure and selection of the appropriate copula. Therefore, to make the case study more representative of real world problems, the sensitivity analysis investigates the effect of adding random noise to the simulated training dataset and its impact on copula selection and parameter estimation. As discussed in Section 8.4.3.2, the first step to generate the training dataset for this case study was to simulate multivariate random vector $\bar{u}$ with marginals uniformly distributed on [0 1] from a *t*-copula. To conduct the sensitivity analysis, 11 experiments were performed by adding a Gaussian noise with mean zero and different standard deviations ranging from 0 to 1 to

the simulated copula random numbers. To investigate the effect of sample size, all the experiments were repeated for two sample sizes of 10,000 and 50,000.

Figures 8.11 and 8.12 show the effect of adding noise on the performance of the copula selection and parameter estimation, respectively, for the local copula between p_{IE} and p_{SB1}. The graphs for the local copulas between other nodes are not included here as they resulted in the same conclusion. As shown in Figure 8.11, for the sample size of 10,000, the AIC approach has selected the *t*-copula as the best fitting approach for the values of noise standard deviation less than 0.5. This shows the robustness of using Akaike weights for copula selection as the value of 0.5 for the noise standard deviation represents a relatively significant noise in real world scenarios since both noise standard deviation and copula simulated random numbers have values between 0 and 1. For higher values of noise standard deviation, the Gumbel copula is selected as the best fitting copula, which is consistent with Table 8.7 where Gumbel has been identified as the second competing copula. As can be seen in Figure 8.11, the robustness of the copula selection method improves by increasing the sample size. These conclusions are also consistent with the observations by Burnham and Anderson [39] and Hashemi et al. [1] on the advantages of using Akaike weights for copula selection.

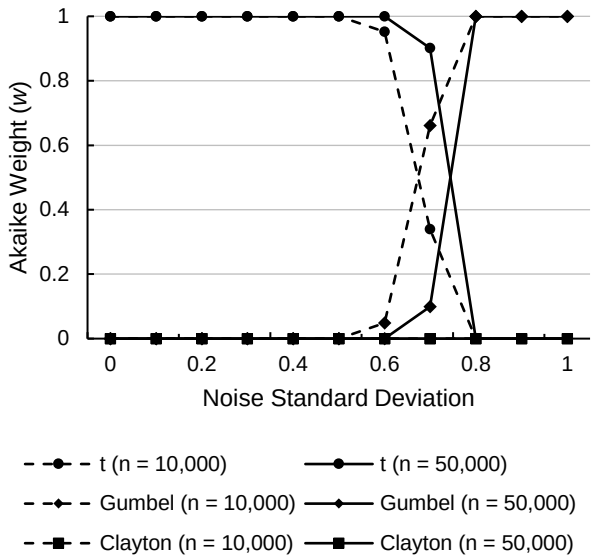


Figure 8.11. The effect of noise standard deviation on copula selection for the local copula for p_{IE} and p_{SB1}. The value of Akaike weight w = 1 represents the best fitting copula model

Figure 8.12 shows the effect of adding noise to the simulated training dataset on parameter estimation using the maximum likelihood (ML) approach presented in Appendix 8.A.2. From Figure 8.12, it can be seen that, compared to copula selection, random noise can have a more negative impact on parameter estimation. Increasing the sample size has also not been as helpful as it was in the case for copula selection. Up to the noise standard deviation of 0.3, noise has had a minor impact on the estimated parameter for the *t*-copula, which is the true model for this case study. However, the performance of the parameter estimation has decreased relatively sharply for a noise standard deviation of more than 0.3. Although a noise standard deviation of 0.3 is still a considerable amount of noise, below which the methodology performance has been reasonably acceptable, another sensitivity analysis is conducted in the following paragraph to investigate the effect of error in parameter estimation in estimated probability values.

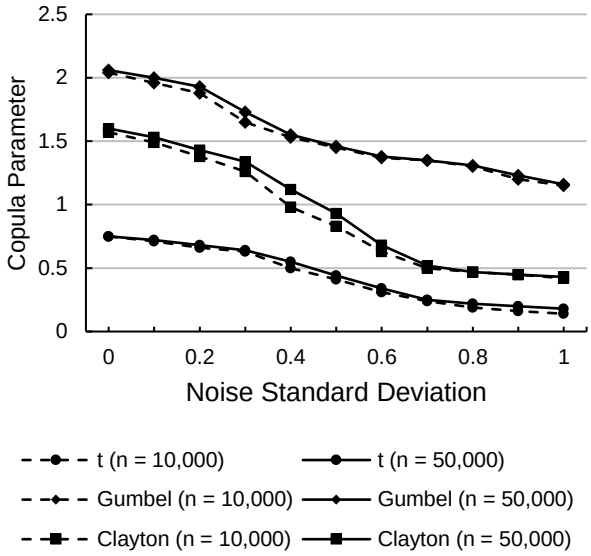


Figure 8.12. The effect of noise standard deviation on copula parameter estimation for the local copula for p_{IE} and p_{SB1}

Figure 8.13 shows the results of the sensitivity analysis that investigates the effect of error in parameter estimation on estimated blowout probability values. As can be seen in Figure 8.13, for the sample size 10,000, -50% and +50% errors in estimated parameters have resulted in about 8% and 13% error in probability estimation, respectively. The error in parameter estimation in the range of -30% to +25% has only produced an error in estimated probability in the range of -5% to +5%. Increasing the sample size has also resulted in better probability estimation. An additional conclusion from Figure 8.13 is that overestimation of the copula parameter is more dangerous than underestimation, as the former has resulted in overestimation of the probability values.

Overall, once the true copula has been selected, which is shown to be of minor concern due to the robustness of the proposed AIC approach, the potential error in parameter estimation due to random noise is of lesser concern. Noise filtering, increased sample size and use of

reliable datasets are among the potential approaches to ensure a reliable parameter estimation.

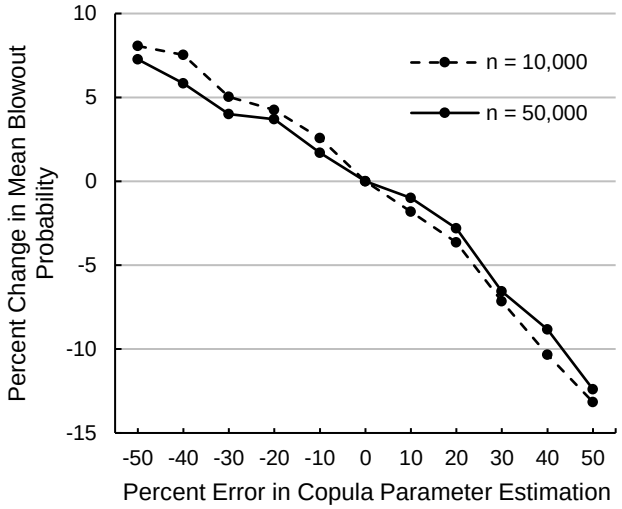


Figure 8.13. Effect of parameter estimation error on probability estimation

8.4.6. Discussion and Further Work

As demonstrated in this case study, the integration of copulas and BN provides safety practitioners with several advantages: (1) it offers control over the form of the univariate marginal distributions; (2) it allows mixing and matching local copulas to represent unknown complex dependence structures; (3) it can accommodate noisy process variables; (4) it can capture hidden relationships among process variables; (5) it features faster structure learning compared to BNs; (6) unlike BNs, CBN does not need discretization of continuous variables, resulting in a significant reduction of information loss and computational cost; and (7) it can capture both causality and dependency interrelationships among random variables.

These advantages make the CBN model highly flexible for different applications where BN analysis cannot be effectively applied, ranging from data-intensive problems like process fault diagnosis to data-scarce scenarios such as risk assessment of new installations or remote operations. The CBN model can be applied for different purposes such as alarm design, safety systems-related decision-making, and risk assessment of multivariate processes.

Like any other modelling technique, the CBN approach presented in this work has some shortcomings, which open new areas of research to further develop its application in process industries. Firstly, the selection of local copulas and associated parameters may be seen as a significant computational burden. It was shown in this case study that the proposed ML method and AIC approach can identify the true dependence structure among variables without significant computational difficulty. However, the AIC attempts only to select the best model from the candidate models available; if a better model exists, but is not offered as a candidate, then the AIC cannot be expected to identify this new model. Therefore, when using the AIC approach, it is strongly recommended to choose all possible copula functions as candidate models. An alternate solution for selection of a copula from among a pre-defined set of copula candidates is the application of sample-based empirical copulas. Genest and Favre [41] believe that the empirical copula "is the most judicious representation of the copula *C* that one could hope for", which, for a bivariate case, is formally defined by:

$$C_n(u,v) = \frac{1}{2}\sum_{i=1}^{n} 1\left(\frac{r_i^u}{n+1} \le u, \frac{r_i^v}{n+1} \le v\right), \quad u,v \in [0,1],$$

where $1(\cdot)$ denotes the indicator function, n is the sample size, and r_i^u and r_i^v stand for the rank of each observation [41].

Another potential challenge for application of copulas is that, when using any copula-based approach, the dependence structure is preserved under strictly increasing (monotonic) relationships among variables [2]. Generally, this is not a restrictive assumption when analysing a network of safety barriers and potential consequences as, intuitively, we do not expect the failure probability of safety barrier SB_i to decrease when the failure probability of SB_{i-1} increases. However, when quantifying the strength of dependence between random variables with nonmonotonic relationships, the monotization transformation technique proposed by Mohseni Ahooyi et al. [2] can be integrated with the presented CBN model to overcome this shortcoming.

The temporal sequence of events does not matter in the presented application of the CBN approach in this work. It is also assumed in this case study that the dependence structure among network nodes remains constant during the drilling operations and at different drilling depths. However, extending the CBN model into a dynamic model, which is an interesting topic for further study, facilitates the incorporation of shifting operating conditions and modes that may change the root nodes, fault scenarios and/or the dependence structure among network nodes. Development of the dynamic CBN to work in the temporal domain will also allow representation and reasoning of the dynamics of complex structured distributions. For a recent attempt to formulate Dynamic Copula Bayesian Network (DCBN) models, an interested reader can refer to Eban et al. [33].

8.5. Conclusions

The concept of copula Bayesian networks (CBN) is applied for process safety analysis by integrating copulas and BN analysis. While copulas capture stochastic dependencies, BN analysis extracts the potential causality and the mutual dependency, also referred to as topological dependencies, among process variables. Thus, the multivariate process safety analysis tool, resulting from the combination of BN analysis and copula functions, provides an intuitively compelling framework for modeling causal relationships among (potentially) highly correlated variables with any level of dependence complexities. The CBN model uses a novel re-parameterization of the conditional densities using copulas. Consequently, the CBN model addresses several shortcomings of traditional BNs. Methods based on maximum likelihood evaluation and Akaike's information criterion (AIC) are used for learning the structure of the CBN model.

The potential broad range of applications for the presented CBN model is demonstrated using a managed pressure drilling case study. The sensitivity analysis conducted for the case study showed that, despite the sensitivity of the model to the choice of copula, the presented AIC approach is capable of selecting the best fitting copula from among a list of candidate copulas, even for datasets with relatively significant random noise. The case study also highlighted the CBN model's relatively low sensitivity to error in copula parameter estimation. In a future study, the model is to be combined with the multivariate operational loss modelling to develop a novel multivariate risk-based process safety analysis tool.

Appendix 8.A. CBN Structure Learning

8.A.1. Estimating Univariate Marginals

Selection of the marginal distributions is not constrained using the proposed CBN approach. Thus, any parametric or non-parametric distribution can be used to represent the marginal distribution of the network nodes. For example, the standard normal kernel-based density estimation can be used to provide an accurate and robust estimation of marginals based on the available data. The non-parametric marginal distributions provide an extremely accurate estimate of the univariate distribution, thus boosting the ability of the CBN model to capture the overall joint distribution [33].

8.A.2. Parameter Estimation of Local Copulas

The maximum likelihood (ML) evaluation method [1] is used here to estimate the copula parameters. Let δ_i be the local copula ratio $R_{c_i}^{\boldsymbol{\delta}_i}(\cdot)$ parameter to be estimated. Given a complete training dataset D of M instances observed from the distribution $f_X(\boldsymbol{x})$ in Equation (8.2) where all of the variables $\boldsymbol{X}$ are observed in each instance, the resulting log-likelihood function can be represented by:

$$\ell(\delta_i) = \sum_{m=1}^{M}\sum_i \log f\left(x_i[m]\right) + \sum_{m=1}^{M}\sum_i \log R_{c_i}^{\delta_i}[m] \tag{8.A1}$$

where $R_{c_i}^{\delta_i}[m]$ is a shorthand for the value that the copula ratio $R_{c_i}^{\boldsymbol{\delta}_i}(\cdot)$ takes in the m'th instance. Then, the copula parameter δ_i is estimated using the ML estimator by maximizing the log-likelihood function of the local copula ratio:

$$\hat{\delta}_i = \arg\max \ell\left(\delta_i\right). \tag{8.A2}$$

The uniform marginals $F\left(\cdot\right)$ required to estimate $R_{c_i}^{\delta_i}\left(\cdot\right)$ in Equation (8.A1) can be estimated using rescaled versions of their empirical counterparts [41], as follows:

$$F_i\left(x\right) = \frac{1}{n+1}\sum_{m=1}^{M} 1\left(X_{im} \le x_i\right), i = 1,...,d\,. \tag{8.A3}$$

Alternatively, the rank-based estimates can also be used to estimate the uniform marginal [1].

8.A.3. Model Selection for Local Copulas

Having estimated the parameters of the candidate copulas, another challenge is to discriminate among competing models. The copula selection method based on Akaike's information criterion (AIC) [1] is used in this work as a formal approach to rank competing copulas for applications in process systems. The AIC approach, with a fundamental link to information theory, uses an empirical log-likelihood function to estimate the relative expected "information" lost, referred to as the Kullback-Leibler (KL) distance, when a candidate model is used to approximate the true (real) model [39]. Using the AIC approach, the model with the smallest estimated AIC value represents the best fitting model.

The AIC of each copula model is estimated using the corresponding value of pseudo log-likelihoods:

$$AIC = -2\ell\left(\delta_i\right) + 2P \tag{8.A4}$$

where P is the number of estimable parameters. It should be noted that the AIC is only comparative relative to other AIC values in the model set [39]. Therefore, the best model is determined by examining its relative distance from the ‘‘true’’ model through computation of the AIC differences, $\Delta_a = AIC_a - AIC_{\min}$, over all candidate models. The smaller Δ_a is, the more likely it is that the adjusted model is the best model. Better interpretation could also be achieved with the Akaike weights [39]:

$$w_a = \frac{\exp(-0.5\Delta_a)}{\sum_{r=1}^{R}\exp(-0.5\Delta_r)}. \qquad (8.A5)$$

The weight w_a is the evidence that model a is the best model, given the data and set of R candidate models. The w_a depends on the entire set; therefore, if a model is added or dropped during a post analysis, the w_a must be recomputed for all the models in the newly defined set. Therefore, it is important to consider all potentially applicable copulas when using the AIC method to rank copulas.

8.6. References

[1] Hashemi SJ, Ahmed S, Khan FI. Operational loss modelling for process facilities using multivariate loss functions. Chem Eng Res Des 2015;104:333–45.

[2] Mohseni Ahooyi T, Arbogast JE, Soroush M. Applications of the Rolling Pin Method. 1. An Efficient Alternative to Bayesian Network Modeling and Inference. Ind Eng Chem Res 2014;54:4316–25. doi:10.1021/ie503585m.

[3] Yu H, Khan F, Garaniya V. A probabilistic multivariate method for fault diagnosis of industrial processes. Chem Eng Res Des 2015;104:306–18. doi:10.1016/j.cherd.2015.08.026.

[4] Oktem UG, Seider WD, Soroush M, Pariyani A. Improve Process Safety with Near-Miss Analysis. Chem Eng Prog 2013;May:20–7.

[5] Elidan G. Copula Bayesian Networks. Adv. neural Inf. Process. Syst. (NIPS 2010), Vancouver: 2010, p. 559–67.

[6] Nelsen RB. An Introduction to Copulas. 2nd ed. New-York: Springer–Verlag; 2006.

[7] Yang F, Duan P, Shah SL, Chen T. Capturing Connectivity and Causality in Complex Industrial Processes. Heidelberg: Springer Science & Business Media; 2014. doi:10.1007/978-3-319-05380-6.

[8] Hashemi SJ, Ahmed S, Khan FI. Correlation and Dependency in Multivariate Process Risk Assessment. 9th IFAC Symp Fault Detect Superv Saf Tech Process 2015:1339–44.

[9] Klaus M. Multivariate Dependence Modeling using Copulas (Master's thesis). Charles University in Prague, 2012.

[10] Pearl J. Fusion, propagation, and structuring in belief networks. Artif Intell 1986;29:241–88. doi:10.1016/0004-3702(86)90072-X.

[11] Mittnik S, Starobinskaya I. Modeling dependencies in operational risk with hybrid Bayesian networks. Methodol Comput Appl Probab 2010;12:379–90. doi:10.1007/s11009-007-9066-y.

[12] Bauer AX. Pair-copula constructions for non-Gaussian Bayesian networks (Doctoral dissertation). Technische Universitat Munchen, 2013.

[13] Pariyani A, Seider WD, Oktem UG, Soroush M. Dynamic Risk Analysis Using Alarm Databases to Improve Process Safety and Product Quality: Part II - Bayesian Analysis. AIChE J 2012;58:826–41. doi:10.1002/aic.12642.

[14] Khakzad N, Khan F, Paltrinieri N. On the application of near accident data to risk analysis of major accidents. Reliab Eng Syst Saf 2014;126:116–25. doi:10.1016/j.ress.2014.01.015.

[15] Yu H, Khan F, Garaniya V. Nonlinear Gaussian Belief Network Based Fault Diagnosis for Industrial Processes. J Process Control 2015:In press. doi:10.1016/j.jprocont.2015.09.004.

[16] Kjaerullf UB, Madsen AL. Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis. New York: Springer; 2008.

[17] Bobbio A, Portinale L, Minichino M, Ciancamerla E. Improving the analysis of dependable systems by mapping Fault Trees into Bayesian Networks. Reliab Eng Syst Saf 2001;71:249–60. doi:10.1016/S0951-8320(00)00077-6.

[18] Sklar A. Fonctions de répartition à n dimensions et leurs marges. Publ Inst Stat Univ Paris 1959:229–31.

[19] Hashemi SJ, Ahmed S, Khan F. Correlation and Dependency in Multivariate Process Risk Assessment. IFAC SAFEPROCESS 2015 9th IFAC Symp. Fault Detect. Superv. Saf. Tech. Process. IFAC-PapersOnLine, vol. 48, Paris: Elsevier Ltd.; 2015, p. 1339–44. doi:10.1016/j.ifacol.2015.09.711.

[20] Hashemi SJ, Ahmed S, Khan FI. Loss scenario analysis and loss aggregation for process facilities. Chem Eng Sci 2015;128:119–29. doi:10.1016/j.ces.2015.01.061.

[21] Hashemi SJ, Ahmed S, Khan F. Probabilistic modelling of business interruption and reputational losses for process facilities. Process Saf Prog 2015;34:373–82. doi:10.1002/prs.11753.

[22] Mohseni Ahooyi T, Arbogast E, Soroush M. Rolling Pin Method: Efficient General Method of Joint Probability Modeling. Ind Eng Chem Res 2014;53:20191–203. doi:10.1021/ie503584q.

[23] Kirshner S. Learning with Tree-Averaged Densities and Distributions. Adv. Neural Inf. Process. Syst. 20, vol. 20, Vancouver: 2007, p. 761–8.

[24] Aas K, Czado C, Frigessi A, Bakken H. Pair-copula constructions of multiple dependence. Insur Math Econ 2009;44:182–98. doi:10.1016/j.insmatheco.2007.02.001.

[25] Bedford T, Cooke RM. Vines: A new graphical model for dependent random variables. Ann Stat 2002;30:1031–68.

[26] Elidan G. Copulas in Machine Learning. In: Jaworski P, Durante F, Härdle WK, editors. Copulae Math. Quant. Financ., Berlin: Springer; 2013, p. 39–60. doi:10.1007/978-3-642-35407-6.

[27] Hanea AM, Kurowicka D, Cooke RM. Hybrid method for quantifying and analyzing Bayesian belief nets. Qual Reliab Eng Int 2006;22:709–29. doi:10.1002/qre.808.

[28] Hanea AM, Kurowicka D. Mixed non-parametric continuous and discrete Bayesian belief nets. In: T. Bedford, J. Quigley, L.Walls, B. Alkali, A. Daneshkhah and GH, editor., 2008, p. 9–16.

[29] Hanea AM, Kurowicka D, Cooke RM, Ababei D a. Mining and visualising ordinal data with non-parametric continuous BBNs. Comput Stat Data Anal 2010;54:668–87. doi:10.1016/j.csda.2008.09.032.

[30] Silva R, Gramacy R. MCMC methods for Bayesian mixtures of copulas. Proc. AI Stat. Conf., Clearwater Beach: 2009, p. 512–9.

[31] Ahmed S, Dalpatadu P, Khan F. Conceptual framework for an event-based plant alarm system 2014:491–6.

[32] Khan F, Abbasi S. A criterion for developing credible accident scenarios for risk assessment. J Loss Prev Process Ind 2002;15:467–75. doi:10.1016/S0950-4230(02)00050-5.

[33] Eban E, Rothschild G, Mizrahi A, Nelken I, Lidan G. Dynamic Copula Networks for Modeling Real-valued Time Series. 16th Int. Con- ference Artif. Intell. Stat., vol. 31, Scottsdale, AZ: 2013, p. 247–55.

[34] Elidan G. Lightning-speed Structure Learning of Nonlinear Continuous Networks. Int. Conf. Artif. Intell. Stat., La Palma: 2012, p. 355–63.

[35] Tenzer Y, Elidan G. Speedy Model Selection (SMS) for Copula Models. Uncertain Artif Intell 2013:1–10.

[36] Bao H, Khan F, Iqbal T, Chang Y. Risk-Based Fault Diagnosis and Safety Management for Process Systems. Process Saf Prog 2011;30:6–17. doi:10.1002/prs.10421.

[37] Abimbola M, Khan F, Khakzad N. Dynamic safety risk analysis of offshore drilling. J Loss Prev Process Ind 2014;30:74–85. doi:10.1016/j.jlp.2014.05.002.

[38] Abimbola M, Khan F, Khakzad N, Butt S. Safety and risk analysis of managed pressure drilling operation using Bayesian network. Saf Sci 2015;76:133–44. doi:10.1016/j.ssci.2015.01.010.

[39] Burnham KP, Anderson DR. Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach. 2nd ed. Secaucus, NJ, USA: Springer; 2002.

[40] Izadi I, Shah SL, Shook DS, Kondaveeti SR, Chen T. A Framework for Optimal Design of Alarm Systems. IFAC Proc Vol 2009;42:651–6. doi:10.3182/20090630-4-ES-2003.00108.

[41] Genest C, Favre A-C. Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask. J Hydrol Eng 2007:347–68.

9. SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

9.1. Summary

The evolution of risk-based approaches and the main contributions in the area of dynamic risk assessment are investigated in this thesis. Comparing the strengths and limitations of different presented dynamic risk assessment methods, the current technological challenges to the development of an efficient and practical dynamic risk assessment approach are identified. Loss functions and Copula Bayesian Networks (CBN) are used to address the identified challenges for dynamic estimation of risk elements, which are system loss and probability of loss occurrence.

Loss functions are used to design an operational risk-based warning system in this work. This is a paradigm shift that will benefit the process industry by continuously improving process safety through proactive loss minimization. Instead of relying on the safety level which has been designed during the system design stage, the utilization of the proposed approach integrates safety improvement into daily activities through loss minimization.

This thesis also demonstrates the flexibility and strength of copula-based approaches in modelling dependence among random variables. Additionally, copulas are used to join univariate marginal loss functions and develop multivariate loss functions. The practical application of copula-based loss aggregation and the importance of considering the correlation and dependency in risk assessment of multivariate processes are highlighted using several case studies.

The concept of Copula Bayesian Networks (CBN) is introduced in this work for process safety analysis by integrating copulas and BN analysis. While copulas capture stochastic dependencies, BN analysis extracts the potential causality and the mutual dependency, also referred to as topological dependencies, among process variables. Thus, the multivariate process safety analysis tool resulting from the combination of BN analysis and copula functions provides an intuitively compelling framework for modeling causal relationships among (potentially) highly correlated variables with any level of dependence complexities. From an application point of view, the aforementioned methods have been effectively applied to model a wide range of complex accident scenarios, from different process systems to offshore drilling operations.

9.2. Conclusions

Considering the importance of making decisions based on real-time risk, there have been efforts to make risk assessment methods dynamically adaptable. Having a dynamic operational risk assessment tool provides a real-time metric to measure and monitor process safety and quality performance. This thesis provides a practical infrastructure to facilitate real-time evaluation of risk elements, which are loss and its probability. This outcome facilitates real-time estimation of risk and its application in risk-based safety management of process facilities. The specific conclusions of this thesis are as follows:

9.2.1. Application of Loss Functions in Process Safety

Loss functions are the perception of process loss due to process variations. Inverted probability loss functions provide more flexibility to model system loss compared to the traditional quadratic loss function. However, choice of the exact form of loss function should be based on the process behaviour and availability of loss data. Use of loss functions helps to continuously update the operational loss according to the current state of the process. Instead of relying on the safety level considered at the design stage, the estimated loss profile, in conjunction with the loss probability, will enable operators to make informed decisions based on the real- time operational risk.

9.2.2. Development of Multivariate Loss Functions in Process Safety

In practical applications a typical process system possesses multiple key characteristics related to product quality and process safety. The proposed multivariate loss function approach using copulas allows for the selection of any type of inverted probability loss function for the marginal losses, irrespective of their dependence structure. The case study results show a significantly improved representation of the process loss behaviour when using the presented copula-based multivariate loss function methodology instead of the classical multivariate inverted normal loss function approach.

9.2.3. Dependency in Multivariate Process Risk Assessment

This thesis discusses the importance of considering dependency among frequencies and loss severities of risk factors. The findings from the case studies highlight the fact that

ignorance or simplification of dependence structures among risk factors, or improper estimation of dependence structures, could result in significant over- or under-estimation of the overall risk. Given the complexity of relationships among process variables, dependence can be quantified in more sophisticated ways than merely through numeric coefficients such as the Pearson and rank correlation coefficients. Copula functions, sometimes referred to as "dependency functions", contain all of the dependence information among random variables. The case study results demonstrated the flexibility and strength of copula-based approaches in modelling the dependence among random variables. Using copulas, the dependence pattern of the random variables and their individual behaviours, and more precisely, their marginal probability distributions, can be studied separately.

9.2.4. Estimation of Business Losses

In this study, models are proposed to assess the business loss elements, including business interruption loss and reputational loss. Although these two elements are closely correlated, they are modeled separately in this work, as different approaches are required to model each factor. The business interruption loss occurs mainly due to a gap in production, whereas the reputational loss occurs primarily due to the organization's damaged trustworthiness in the marketplace after a major incident, with media coverage, takes place. The case study results show that ignorance of reputational loss and business interruption loss elements can cause significant misestimation of the overall loss.

9.2.5. Development of Copula Bayesian Networks (CBN)

This thesis shows that, despite its broad application in the literature, Bayesian analysis for dynamic probability estimation has several shortcomings such as lack of control over the form of marginal probability distributions and consideration of a simplified dependence structure. The concept of copula Bayesian networks (CBN) is proposed to address the shortcomings of traditional Bayesian analysis. The multivariate process safety analysis tool, resulting from the combination of BN analysis and copula functions, provides an intuitively compelling framework for modeling causal relationships among (potentially) highly correlated variables with any level of dependence complexities.

The selection of local copulas and associated parameters may be seen as a significant computational burden for copula-based approaches. However, it was shown through different case studies that the proposed maximum likelihood (ML) method and Akaike's information criterion (AIC) approach can identify the true dependence structure among variables without significant computational difficulty. Sensitivity analysis showed the robustness of the ML and AIC methods to process noise and sampling variability.

Overall, this PhD thesis provides new methods, insights, definitions, and guidance that:

- Improve the understanding of how to monitor and model more realistically the process risk elements, which are loss and its probability;
- lead to improved risk-informed decision-making at early stages of system failure;
- improve safety and productivity in process operations through dynamic risk evaluation; and
- assist in making zero-accident culture a reality.

9.3. Recommendations

The present work attempts to introduce new concepts and also overcome the limitations of existing techniques in the field of dynamic risk analysis and safety assessment of process industries. This study, however, can be extended further as suggested below to address the main limitations of the work, as identified in the following sections.

9.3.1. Development of Empirical Copulas

It was shown that the proposed ML method and AIC approach can identify the true dependence structure among variables without significant computational difficulty. However, the AIC attempts only to select the best model from the candidate models available; if a better model exists, but is not offered as a candidate, then the AIC cannot be expected to identify this new model. Therefore, when using the AIC approach, it is strongly recommended to choose all possible copula functions as candidate models. An alternate solution for selection of a copula from among a pre-defined set of copula candidates is the application of sample-based empirical copulas, which can be a subject for further study.

9.3.2. Development of Dynamic Copula Bayesian Network (DCBN) Models

In application of CBN models, it is assumed that the dependence structure among network nodes remains constant over time. Extending the CBN model to become a dynamic model is an interesting topic for further study to facilitate the incorporation of shifting operating conditions and modes that may change the root nodes, fault scenarios and/or the dependence structure among network nodes. The development of the dynamic CBN to

work in the temporal domain will also allow representation and reasoning of the dynamics of complex structured distributions. This will also address the shortcoming of CBN models to allow for feedback events, which consistently exist in process variables.

9.3.3. Development of Data Gathering Methodologies

Most of the proposed approaches in this study demand a high amount of quality data which are often difficult to obtain, particularly for remote operations such as offshore and marine facilities. Choosing appropriate data that best represent the conditions in a given process system is challenging. To tackle this challenge, potential sources of information and data include:

- Expert experience and knowledge
- Data and information shared across industries that have operations in similar environments

The development of advanced data acquisition systems for quantitative risk assessment, as well as development of methodologies for recording and analyzing process-related near misses and unsafe conditions, could be subjects for further studies to systematically gather and share information.

9.3.4. Integration of Dynamic Multivariate Loss and Probability Estimation Methods

The development of dynamic multivariate loss and probability estimation methods is studied separately in this work to enable the elaboration of details of each methodology. The development of a dynamic multivariate risk assessment method can be a subject to

further develop this research by integration of the proposed dynamic multivariate loss and probability estimation methods.

9.3.5. Development of Commercial Tools

MATLAB® codes are used in this thesis for the development and implementation of the proposed models. However, there is a need to develop a commercial and user-friendly software tool for implementation of dynamic multivariate risk analysis. The developed software tool should be compatible with current data acquisition and control systems in the oil and gas industry to facilitate its application for real-time risk control and its application for activation of process control and safety devices.

9.3.6. Development of Dynamic Risk Management Tool

It has been shown in numerous studies that several major losses in the oil and gas industry might have been prevented if a dynamic risk approach like the one presented in this work was integrated into the management framework. However, the implementation of a dynamic risk assessment approach could be a complex, resource-demanding process. Therefore, as suggested by Paltrinieri et al.[8], "a strong safety culture for monitoring and recording process performances and incidents is needed, and a robust decision-making process should be introduced". The integration of proposed loss and probability assessment techniques with management systems, along with a strong safety culture promoting

[8] Paltrinieri, N., Khan, F., Cozzani, V., 2014. Coupling of advanced techniques for dynamic risk management. J. Risk Res. 9877, 1–21. doi:10.1080/13669877.2014.919515.

continuous improvement, are important subjects for future study. The resulting dynamic risk management tool would assist managers and decision-makers in making an adequate and timely investment in safety measures.

9.3.7. Conducting an Approximate Uncertainty Modeling

Probability distributions are used in this work to model multiple uncertainties. Uncertainty associated with the selection of proper probability distributions and the estimation of probability distribution parameters can significantly impact the accuracy of risk assessment. Uncertainty analysis investigations have been conducted in different parts of this thesis to address this challenge. However, a more formal uncertainty modelling study by separating the epistemic and aleatory uncertainties is recommended to ensure consideration of all sources of uncertainty when applying the proposed methods in this thesis. A recent study by Bedford et al.[9] can be used as a guideline to approximate uncertainty modeling in risk analysis.

9.3.8. Practical Application in Process Safety Monitoring

Figure 9.1 shows a simple schematic illustration of how the proposed Dynamic Risk Assessment (DRA) methods can be incorporated into the process safety monitoring system and compares it with the traditional approach. In the traditional safety system design, variable deviations from their predefined threshold limits are monitored to activate control

[9] Bedford, T., Daneshkhah, A., Wilson, K.J., 2016. Approximate Uncertainty Modeling in Risk Analysis with Vine Copulas. Risk Anal. 36, 792–815. doi:10.1111/risa.12471

systems, alarms, and emergency shutdown devices (ESD). However, compared to this conventional single variable-based safety system design approach, the application of the proposed DRA methods ensures continuous updating of the events risk based on new evidence from: (i) monitoring of multiple correlated process variables; (ii) failure and incident histories; and (iii) process and operational changes. Therefore, safety and control limits can be set on the basis of the estimated risk. Moreover, the estimated risk can be used to categorize warnings into alerts and alarms to avoid problems such as alarm flooding.

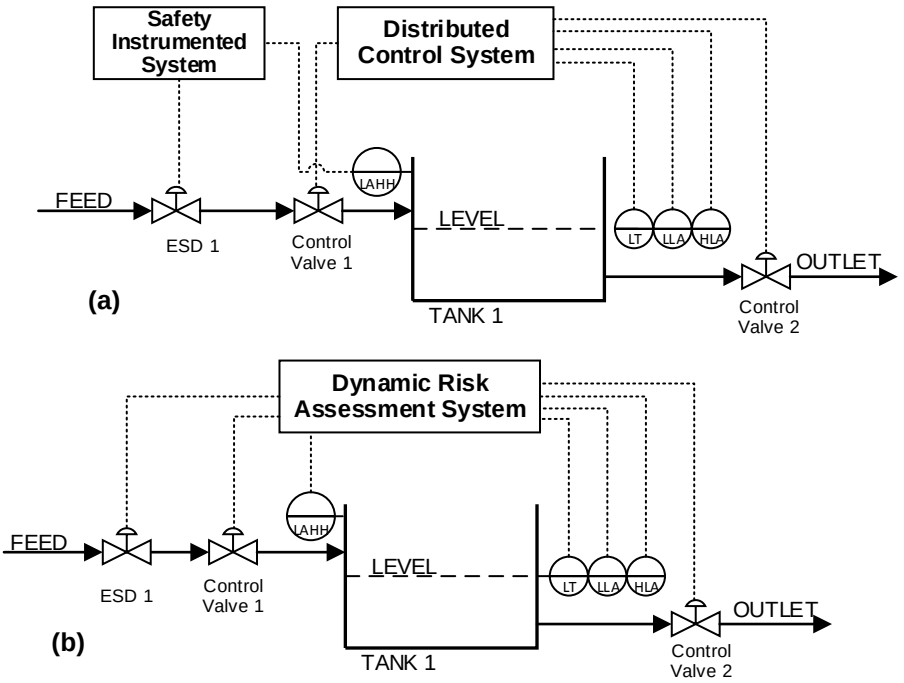


Figure 9.1. Schematic of a tank system (a) Traditional safety and control system design; (b) Incorporation of the proposed Dynamic Risk Assessment methods in process safety monitoring[10]

[10] Khan, F., Hashemi, S.J., Paltrinieri, N., Amyotte, P., Cozzani, V., Reniers, G., 2016. Dynamic risk management: a contemporary approach to process safety management. Curr. Opin. Chem. Eng. 14, 9–17. doi:10.1016/j.coche.2016.07.006