UNIVERSITY OF STRATHCLYDE

DEPARTMENT OF MANAGEMENT SCIENCE

Uncertainty Analysis of Large Risk Assessment Models with Applications to the Railway Safety and Standards Board Safety Risk Model

by

Daosheng Cheng

A thesis presented in fulfilment of the requirements for the degree of Doctor of Philosophy

September 2009

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has lead to the award of a degree.

Signed: Date:

Abstract

Probabilistic risk analysis aims to assess the safety risk of a system so that actions can then be taken to improve safety. Uncertainty however always exists in modelling. For more informed decision making, uncertainty in the outputs of the model must be assessed through uncertainty analysis.

This research focuses on parameter uncertainty of a risk model composed of fault trees and event trees. Research questions include: (1) how to model the subjective uncertainty in the basic events and the consequences; (2) how to propagate the uncertainty in the input parameters through fault trees and event trees to obtain uncertainty in the output.

Structured approaches are developed to elicit the covariance matrix of the basic events and to model dependence among the consequences. To calculate the uncertainty propagation, a model is developed to mimic fault trees and event trees; an analytical solution and a simulation-based method are developed for assessing the uncertainty propagation, which are implemented independently and therefore cross-check each other.

The developments can be used for subjective uncertainty assessment of Fault-tree and Event-tree models. With the developed methods, a reasonable elicitation workload is required to model the subjective uncertainty in the input parameters; the assessments can be monitored during the elicitation process. The methods for assessing the uncertainty in the output can work efficiently for large fault trees and event trees.

Two case studies have been conducted with the Safety Risk Model (SRM) developed by Rail Safety and Standard Board (RSSB), UK. In the two case studies, the developed methods are deployed and experts were confident in making the required assessments. The feasibility of the developments is validated by the case studies.

Acknowledgement

I would like to express my sincere gratitude to my supervisors, Prof. Tim Bedford and Prof. John Quigley. Without their knowledge, experience, and patience, this work would have never been accomplished. Working with Tim and John has been a source of inspiration, a challenge and a pleasure. I would like to thank Dr. J. René van Dorp, Dr. Robert van der Meer and Dr. Tibor Illés for their kind offers to be the examiners and convenor of the examining committee. Based on the comments from the viva, the thesis is improved greatly. I would like to thank Dr. Alireza Daneshkhah, for his valuable comments on this thesis.

The research was funded by the Rail Safety and Standard Board (RSSB), UK. Valuable supports from RSSB made it possible to implement this research. I would like to thank Colin Dennis, Dr. George Bearfield, Kajal Somaiya, Jesse Baker, Dr. David Griffin, and Alan Penter. Special thanks to Chris Harrison for his inspiration, support and cooperation.

For their unlimited support, I am grateful to my father Xinxue Cheng, my mother Yongqin Xie, my sisters and brothers Daoling Cheng, Baowen Qu, Daoquan Cheng, Zhijian Dong, Daojuan Cheng, Anyang Wu, Daoyan Cheng, Ruofeng Li. I would like to give my thanks also to my friends Kun Yang, Jiaxiong Wang, Jianbo Lu, and others both in China and in Scotland, for their encouragement and support.

Thanks are given to all those who helped me finish my research and thesis.

Abbreviations

AHP	Analytic hierarchy process
CDF	Cumulative distribution function
FT/ET	Fault-tree and Event-tree
HET	Hazardous event tree
HSE	Health and Safety Executive
IA	Injury atom
LC	Level crossing
LHS	Latin Hyper-cube sampling
MS	Microsoft
PRA	Probabilistic risk analysis
RSSB	Railway Safety and Standards Board, UK
RV	Road vehicle
SRM	Safety risk model
UA	Uncertainty analysis
UWC	Users worked crossing

Notation

$E(\ )$	mean of a random variable
$r(,)$	correlation coefficient between two random variables
$\mathrm{var}(\ )$	variance of a random variable
p_i	the i^{th} input parameter
μ_i	mean of p_i, i.e. $\mu_i = E(p_i)$
ε_i	residual of the i^{th} parameter p_i in the linear factor model
w_i	coefficient associated with ε_i
σ_i	standard deviation of ε_i, $\sigma_i^2 = \mathrm{var}(\varepsilon_i)$
$X_{i,k}$	an uncertainty factor of p_i
$\lambda_{i,k}$	coefficient associated with $X_{i,k}$
C_k	a factor class of $X_{i,k}$ and $X_{j,k}$ when $r(X_{i,k}, X_{j,k}) \neq 0$
X_k	a common factor of $X_{i,k}$ and $X_{j,k}$ when $X_{i,k} = X_{j,k}$
$L_{E,i}$	set of indices for exclusive factors of p_i
L_C	set of indices for factors belonging to a factor class
L_X	set of indices for common factors
q_i	proportion of $\mathrm{var}(p_i)$ explained by the associated uncertainty factors
f_z	the z^{th} parameter family
$I(f_z)$	set of the invariant factors of the family f_z
$V(f_z)$	set of the variant factors of the family f_z
$\mathbb{R}^{m\times n}$	real matrix composing of *m* rows and *n* columns
E_i	an escalation event
$\mathscr{W}(E_i)$	set of outcomes of E_i
$\omega_{i,j}$	the j^{th} outcome of E_i, $\omega_{i,j} \in \mathscr{W}(E_i)$

S_ℓ	the ℓ^{th} accident sequence
$e(\ell,i)$	outcome of E_i associated with S_ℓ, $e(\ell,i)\in \mathscr{W}(E_i)$
$\pi_\ell(m)$	individual injury probabilities associated with S_ℓ, where $m=1,\cdots,4$ stand for no injury, minor injury, major injury, and fatality respectively
N_ℓ	mean of the number of people exposed to the risk scenario S_ℓ
$c_\ell(m)$	consequence at level *m* associated with S_ℓ
T_j	the j^{th} hazard type
$\mathscr{E}(T_j)$	subset of escalation events that affect T_j
$H_{\ell,j}$	hazard of the type T_j specified by S_ℓ
R_ℓ	rule set, $R_\ell=(H_{\ell,1},\cdots,H_{\ell,j},\cdots,H_{\ell,D})$
$A_{i,j}(m,k)$	hazard $H_{i,j}$'s transition probability from *m* level to *k* level
$A_{i,j}(m)$	transition probability from *m*-level injury to all higher levels of injuries
$\varphi_{i,j}(m,k)$	mean of $A_{i,j}(m,k)$
A_j	injury atom associated with T_j
γ_j	mean of A_j

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Outline

The goal of this research is to develop methods to support the assessment of subjective uncertainty in the output of a risk assessment model composed of fault trees and event trees. In this chapter, we start with an introduction to risk and probabilistic risk analysis. We continue to an introduction to fault trees and event trees as they are widely used in probabilistic risk analysis. We then continue to describe how decisions are made based on the probabilistic risk analysis to reduce risk. We can not however assess with certainty the input parameters of a probabilistic risk analysis model. Consequently we introduce uncertainty and its important role in the risk assessment context. We then identify three research aims forming the research goal. An overview of this thesis is then given at the end of this chapter.

1.2 Probabilistic Risk Analysis

As defined in [HSE(Health & Safety Executive) 2001], a hazard is the potential for harm arising from an intrinsic property or disposition of something to cause detriment. Some hazards are summarized in [Modarres 2006] as:

- Chemical (e.g., toxins, corrosive agents, smoke)
- Biological (e.g., viruses, microbial agents, bio-contaminants)
- Thermal (e.g., explosion, fire)
- Mechanical (e.g., impact from a moving object, explosion)
- Electrical (e.g., electromagnetic fields, electric shock)
- Ionizing radiation (e.g., x-rays, gamma rays)

- Nonionizing radiation (e.g., microwave radiation, cosmic rays)

A hazard therefore is a source of danger. When people are exposed to hazards, they suffer the chance to be adversely affected by the hazards, which is called risk in [HSE(Health & Safety Executive) 2001]. Some people argue that the adverse impact on the unlucky people caught in an accident cannot be completely measured [Adams 2001]. It is however the potential injuries and fatalities that primarily concern us; hence, we usually mean safety risk when we refer to risk. When a system is subject to hazards, the associated risk can be measured qualitatively or quantitatively. In a qualitative way, the risk is assessed on several levels from low to high. In a quantitative way, the risk may be measured as the number of the potential injuries and fatalities over a unit of time.

Probabilistic risk analysis (PRA) is a systematic procedure for assessing the quantitative risk of a system [Bedford and Cooke 2001; Abrahamsson 2002; Jonkman, Van Gelder et al. 2003; Dennis 2006; Modarres 2006]. It is also called quantitative risk analysis or probabilistic safety analysis (PSA). To conduct a PRA, we need to answer three questions as given in [Kaplan and Garrick 1981] as

i. What can go wrong?
ii. How likely is it to happen?
iii. Given that it occurs, what are the consequences?

To answer the first question we need to identify the risk scenarios of the system. The risk scenario is defined at the end of an accident sequence [Bedford and Cooke 2001] Usually the risk scenarios are modelled with event trees that we are going to introduce in the next section. To answer the second question, the frequency of each risk scenario can be estimated. To answer the third question on the above list, the consequences of a risk scenario are defined as the numbers of the injuries sometimes including minor injuries, major injuries and fatalities. We denote a risk scenario as S_i; we denote by f_i and c_i the frequency and the consequence of S_i respectively.

The risk is then defined by a set of triplets $\langle S_i, f_i, c_i \rangle$, $i = 1, \cdots, n$, [Kaplan and Garrick 1981]. Associated with each risk scenario, the risk may be defined as

$$r_i = f_i \times c_i \tag{1.1}$$

and the risk of the system defined as

$$r = \sum_{i=1,n} r_i \tag{1.2}$$

This is the expected value of the consequences.

According to Equations 1.1 and 1.2, probabilistic risk analysis consists of three elementary parts including: (1) identifying the risk scenarios; (2) estimate the frequencies of the scenarios; (3) estimate the consequences of the scenarios. Fault-tree and Event-tree models are the most popular tools for these tasks [Bedford and Cooke 2001; Abrahamsson 2002]. We are going to give an introduction to Fault-tree and Event-tree models in the next section.

1.3 Fault-tree for Event-tree Models

Fault-tree and Event-tree models have been used extensively in PRA [Kumamoto and Henley 1996; Bedford and Cooke 2001; Abrahamsson 2002]. Associated with the system of interest, we can identify an initiating event, or several such events, which has the potential to cause a series of hazards. After the initiating event we can identify a sequence of escalation events, for which the outcomes affect the final consequence. Following a "forward logic", an Event-tree model begins with the initiating and continues with the sequence of escalation events. Each event is represented by a node in the event tree; the outcomes of an event are represented by the branches following the associated nodes. A path from the initiating event through all the escalation events is called accident sequence; the risk scenario is defined at the end of an accident sequence [Bedford and Cooke 2001]. An example of event

tree is extracted from [Kumamoto and Henley 1996]. Shown in Fig. 1.1 is the schematic diagram of a pressure tank system. After the pressure gas in the tank is discharged, the pump is started to recharge the tank. The tank is protected from overpressure initially by a timer that cuts off the current by opening the contact. The system includes two more protections. One is the operator who opens the manual switch when he reads a high reading of the pressure gauge. The other protection is the relief valve that is designed to open automatically when the tank pressure reaches a preset height. With such a system, the failure of the timer is identified as an initiating event; the two extra protections are identified as the escalation events. The event tree of the example is demonstrated in Fig. 1.2, where three risk scenarios are defined.

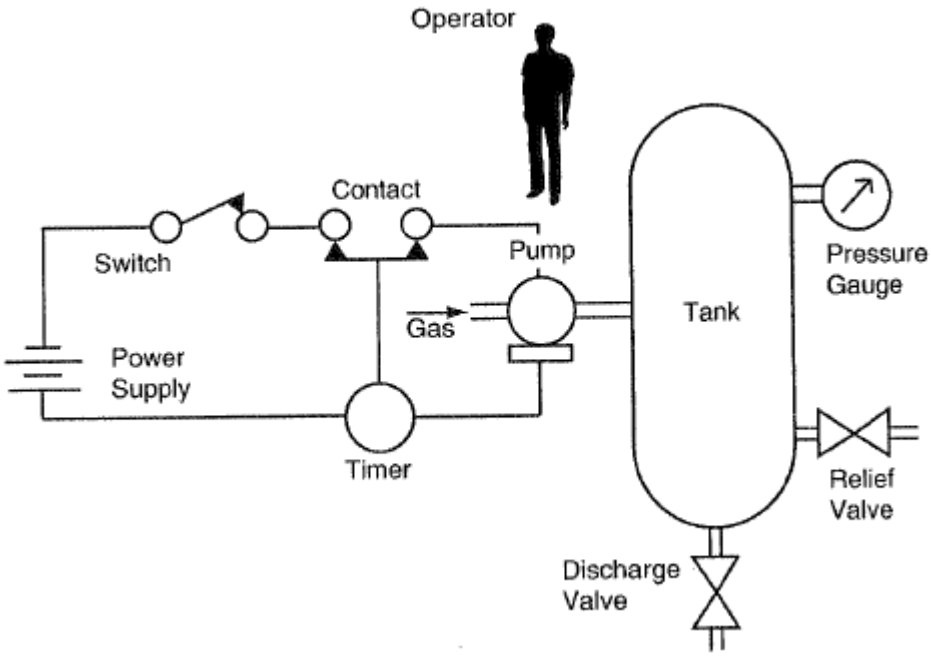


Fig. 1.1 Schematic diagram of pressure tank system as a risk analysis example (reproduced with permission from [Kumamoto and Henley 1996])

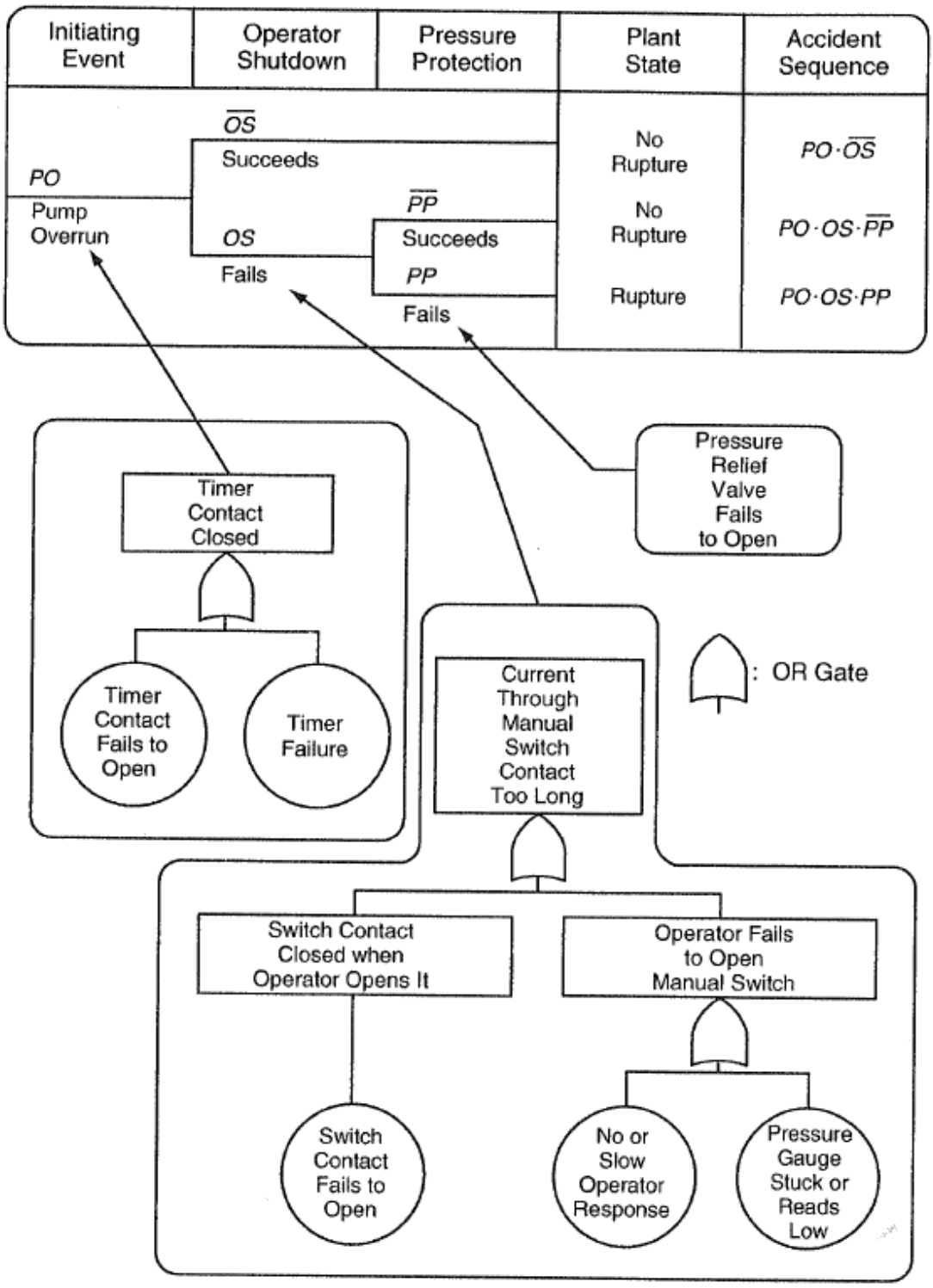


Fig. 1.2 Demonstration of the Fault-tree and Event-tree models (reproduced with permission from [Kumamoto and Henley 1996])

The frequency of the scenario is calculated in terms of the frequency of the initiating event and the probabilities of the outcomes of the escalation events. The frequency and probabilities can be estimated directly for some events. These events can also be broken down to the basic event with fault trees. Following a "backward logic", a fault tree decomposes a particular failure, called the top event, into the basic events that by different combinations lead to the failure. With the probabilities of the basic events estimated, the frequency or the probability of the top events is calculated through the Fault-tree models [Kumamoto and Henley 1996; Bedford and Cooke 2001]. The basic events are sometimes called precursors [Abrahamsson 2002; Dennis 2006]. With a fault tree, we can investigate further the causes of the failures and the connections between the events. As shown in Fig. 1.2, the initiating event pump overrun and the failure of operator shutdown are broken down into the basic events through events trees.

With PRA in conjunction with Fault-tree and Event-tree models, we can quantitatively investigate the risk of a system and the major sources. Actions can be taken accordingly to reduce the risk with the limited resources. We now give an introduction to the *F-N* curve and how it can be made for decision making.

1.4 F-N Curve and Decision Making

Risk cannot be eliminated and resources required to reduce risk are limited. The aim of studying risk is to reduce the risk efficiently with limited resources. The risk as defined in Equation 1.2 offers a measure of the risk level over all the risk scenarios. It however hides the difference between two types of incidents: one of low frequency but high impact consequences; the other one of high frequency but low impact consequences. In fact, people have different attitudes toward these two types of incidents [Health and Safety Executive 1992]. It is therefore of interest to include the frequency and the severity profile in the risk assessment. It can be represented by the *F-N* curve, where N stands for the fatalities in one incident; F stands for the yearly frequency of the incidents causing N or more fatalities. As the *F-N* curves were firstly made by Farmer in 1967, they are also called Farmer curves. The *F-N* curve

can be built based on the frequencies and the consequences of the risk scenarios S_i, $i = 1, \cdots, n$. We at first order the risk scenarios to satisfy $c_{i-1} < c_i$. We can then plot the cumulative frequency $F_i = \sum_{k=i,n} f_k$ against c_i for $i = 1, \cdots, n$. Because both *F* and *N* can range across several orders of magnitude, we usually draw the *F-N* curve on logarithmic scales. For examples, the *F-N* curves are made for road transport, rail yards, and airports etc as shown in Fig. 1.3.

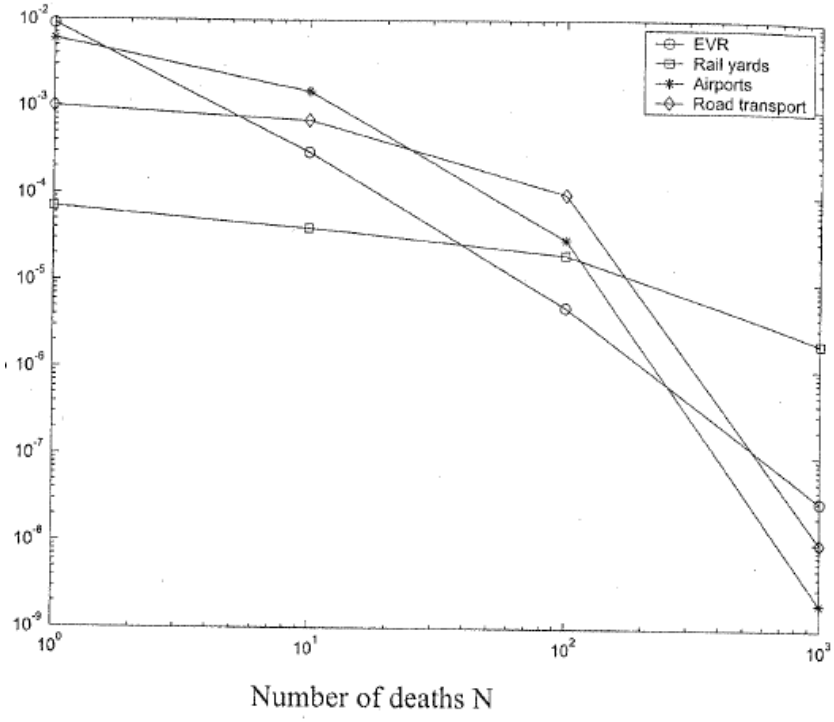


Fig. 1.3 Example of F-N Curves (reproduced with permission from [Bedford and Cooke 2001])

We can also set up the risk tolerability criteria on the *F-N* Curves. As shown in Fig. 1.4, two *F-N* lines partition the positive quadrant into three areas marked as unacceptable, ALARP and acceptable respectively. The acronym ALARP stands for "as low as reasonably possible". If any part of a system's *F-N* curve enters the unacceptable area, the associated risk is regarded as intolerable. Safety actions must be taken to lower the *F-N* curve accordingly. If a system's *F-N* curve is confined to the acceptable area, the associated risk can be regarded as tolerable. In other cases

when a system's *F-N* curve completely or partly lies in the ALARP area, safety actions can be pursued based on the cost-benefit analysis. We need to analyze how much extra safety can be gained with extra money because the resource is always limited.

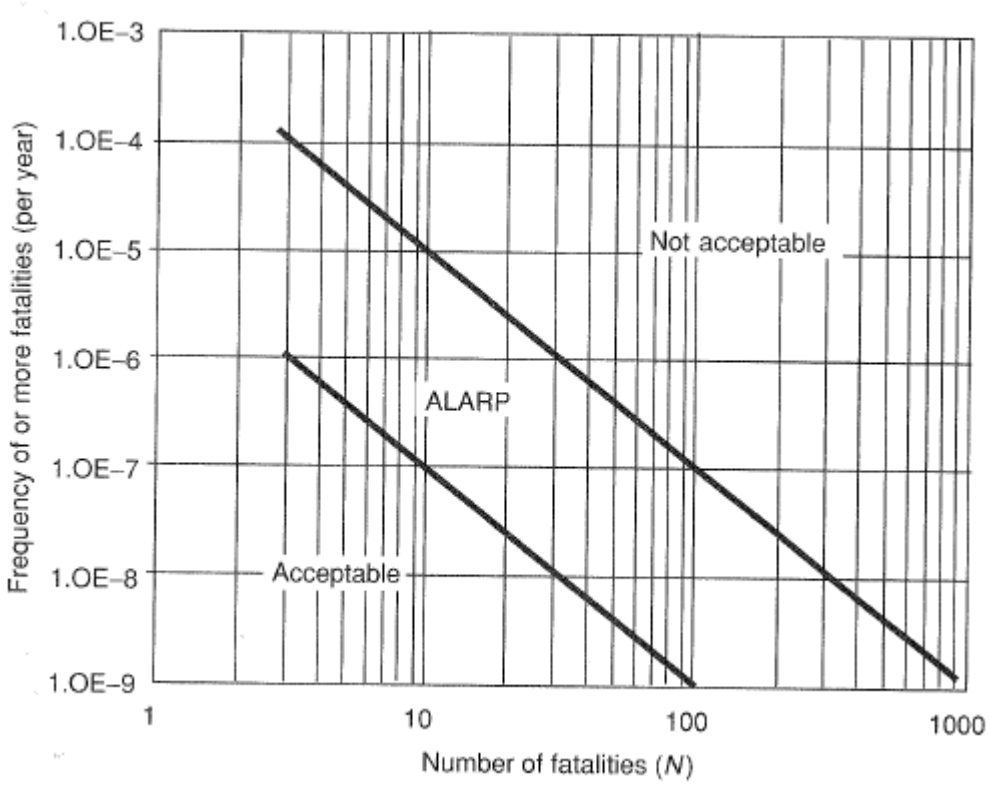


Fig. 1.4 Demonstration of the risk tolerability criteria defined by F-N curves (reproduced with permission from [Haugom, Rikheim et al. 1990])

The *F-N* curves have been used in risk assessment in various contexts in several countries for about three decades [Evans 2003]. In the literature, we can see two risk tolerability criteria that are defined by the F-N curves. One is recommended by HSE [Health and Safety Executive 1992]; the other is recommended by Netherlands Planning Department [Versteeg 1988]. Both criteria are defined by two parallel straight lines. The HSE criterion however has the slope -1; while the Dutch criterion

has the slope -2. Therefore Dutch criterion has stronger aversion to accidents with multiple fatalities.

With the PRA and the criteria on the *F-N* curves, we can judge whether more actions are needed to improve the safety of a given system. However, we cannot assess with certainty the probabilities of the basic events and the consequence associated with each risk scenario. This uncertainty plays an important role in decision making [Morgan and Henrion 1990]. Therefore we need to study the uncertainty in the input parameters and then assess the uncertainty in the output. The next section will introduce uncertainty.

1.5 Uncertainty Analysis

1.5.1 Uncertainty Classes

Typically uncertainty is classified into aleatory uncertainty and epistemic uncertainty. Aleatory uncertainty describes natural variability. Therefore aleatory uncertainty is also called stochastic uncertainty. Aleatory uncertainty can be quantified by measurements and statistical estimations, or by expert judgement. Epistemic uncertainty represents the lack of knowledge. Therefore it is also called knowledge-based uncertainty. Epistemic uncertainty can be quantified by expert opinion [Bedford and Cooke 2001]. In practice, it depends on both the current knowledge and the measures we use to assess uncertainty. The uncertainty of a complex system usually is affected by many factors and therefore is difficult to classify. It can be decomposed into more understandable and manageable sources. The system uncertainty can then be classified into epistemic parts and aleatory parts. This classification is very helpful in practice [Winkler 1996; Bedford and Cooke 2001]. Firstly it makes clear that what kinds of methods can be used to model the uncertainty. Secondly the classification helps design suitable ways to quantify the uncertainty. Thirdly, the classification informs the decision maker about the effects of the epistemic uncertainties on the model output that could be reduced by more learning.

1.5.2 Importance of Modelling the Uncertainty and Dependence of the Input Parameters

To conduct PRA, we need to evaluate the basic events and the consequences associated with the scenarios. In the context of risk analysis, generally there are not sufficient risk occurrences from which we can estimate the values of the input parameters. Consequently these values are mainly elicited from expert judgement [Cooke 1991; Iman and Helton 1991; Bedford and Cooke 2001; Abrahamsson 2002]. Due to the lack of the knowledge, however, we cannot assess with certainty the values of these input parameters. Consequently uncertainty is introduced in implementing the probabilistic risk analysis due to imperfect knowledge of the input parameters [Winkler 1996; Abrahamsson 2002; Lauridsen, Kozine et al. 2002]. This uncertainty belongs to epistemic uncertainty as discussed above. An example of the uncertainty in the PRA input parameters can be found in the project ASSURANCE (ASSessment of Uncertainty in Risk Analysis of Chemical Establishments) [Lauridsen, Kozine et al. 2002]. In this project, seven teams from different European countries were asked to assess the frequencies and the consequences of 11 scenarios related to an ammonia storage facility. As summarized in Table 1.1 and Fig. 1.5, the results show a wide spread for both the frequencies and the consequences.

Table 1.1 Example of the variability in the assessed frequencies of the reference scenarios, reproduced with permission from [Lauridsen, Kozine et al. 2002]

		Partner number							
#	Top Event*	3	4	1	5	7	2	6	Range of deviation
1	Major ammonia leak from 8'' feeding pipe	$2.1\ 10^{-4}$	$5.0\ 10^{-5}$	$9.5\ 10^{-5}$	$1.6\ 10^{-5}$	$2.0\ 10^{-2}$	$7.7\ 10^{-6}$	3	$5.0\ 10^{-6}$ - $2.1\ 10^{-4}$
2	Breakage of 4'' pipe 241P-067-P1349	$3.9\ 10^{-4}$	$1.0\ 10^{-4}$	$2.0\ 10^{-4}$	$5.9\ 10^{-5}$	$7.3\ 10^{-4}$	$4.5\ 10^{-4}$	2	$5.9\ 10^{-5}$ - $7.3\ 10^{-4}$
4	Rupture or disconnection between ammonia ship and unloading arm 241-ME1	$5.8\ 10^{-3}$	$5.0\ 10^{-3}$	$4.8\ 10^{-4}$	$4.1\ 10^{-6}$	$1.0\ 10^{-5}$	$4.8\ 10^{-4}$	4	$4.1\ 10^{-6}$ - $5.8\ 10^{-3}$
7	Rupture of 10'' pipe 241P-089-P1283	$4.0\ 10^{-4}$	$2.0\ 10^{-8}$	$3.9\ 10^{-8}$	$7.0\ 10^{-5}$	$1.7\ 10^{-4}$	-------	2	$2.0\ 10^{-8}$ - $4.0\ 10^{-4}$
7*	Rupture of a ship tank	$5.7\ 10^{-3}$	------	$2.3\ 10^{-7}$	$2.3\ 10^{-6}$	$4.9\ 10^{-6}$	$2.3\ 10^{-7}$	-------	$2.3\ 10^{-7}$ - $5.7\ 10^{-5}$
9	Rupture of cryogenic tank 241-S1	Contained leak: $1.0\ 10^{-6}$ Uncontain leak: $4.0\ 10^{-8}$	------	$5.0\ 10^{-7}$	$5.0\ 10^{-8}$	$5.0\ 10^{-7}$	$1.0\ 10^{-8}$	4	$1.0\ 10^{-8}$ - $1.0\ 10^{-6}$
10	Rupture of 20'' pipe 241P-015-P1284	$9.0\ 10^{-5}$	$1.0\ 10^{-5}$	$7.6\ 10^{-6}$	$8.8\ 10^{-7}$	$9.7\ 10^{-7}$	$1.0\ 10^{-6}$	2	$8.7\ 10^{-7}$ - $9.0\ 10^{-5}$
14	Rupture of one of the ten pressurised tanks	$2.5\ 10^{-6}$	$5.0\ 10^{-7}$	$1.6\ 10^{-6}$	$1.3\ 10^{-3}$	$2.0\ 10^{-6}$	$5.0\ 10^{-7}$	3	$5.0\ 10^{-7}$ - $1.3\ 10^{-5}$
15	Rupture of 4'' pipe on the distribution line of tank 241-V1	$2.3\ 10^{-4}$	$2.0\ 10^{-5}$	$6.0\ 10^{-5}$	$1.1\ 10^{-5}$	$4.9\ 10^{-7}$	$3.4\ 10^{-8}$	2	$3.4\ 10^{-8}$ - $2.3\ 10^{-4}$
17	Rupture or disconnection between ammonia truck and unloading arm	$3.7\ 10^{-3}$	$6.0\ 10^{-5}$	$4.7\ 10^{-6}$	$6.8\ 10^{-5}$	$1.0\ 10^{-6}$	$1.5\ 10^{-7}$	1	$1.5\ 10^{-7}$ - $3.7\ 10^{-3}$
18	Catastrophic rupture of a truck tank	$2.3\ 10^{-7}$	$1.2\ 10^{-7}$	$1.1\ 10^{-8}$	$7.4\ 10^{-9}$	$2.7\ 10^{-8}$	$1.5\ 10^{-9}$	1-2	$1.5\ 10^{-9}$ - $2.3\ 10^{-7}$

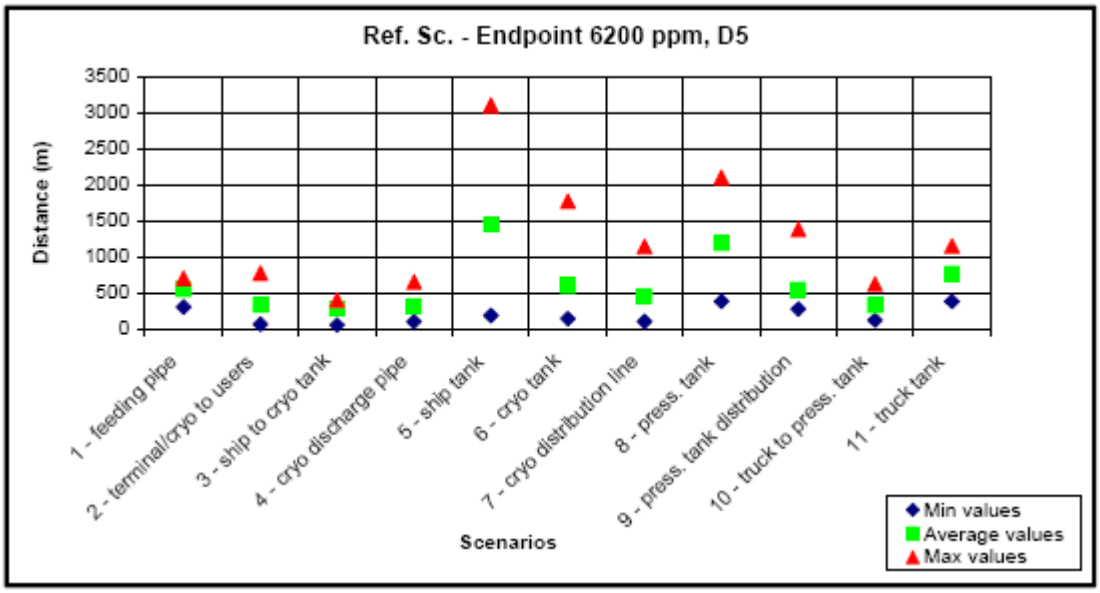


Fig. 1.5 Example of the variability in the consequence assessment of the 11 reference scenarios, reproduced with permission from [Lauridsen, Kozine et al. 2002]

The uncertainty in the input parameters consequently causes the uncertainty in the output of a PRA model. This is demonstrated with the example regarding the project ASSURANCE. Based on the assessments of the input parameters as summarized in Table 1.1 an Fig. 1.5, one *F-N* curve is drawn based on each team's assessments. As shown in Fig. 1.6, the *F-N* curves show a considerable spread. For instance, the frequency of events leading to 100 or more fatalities ranges over two orders of magnitude among the different teams' assessments. Consequently, the risk level could be judged to be tolerable or unacceptable depending on which assessment we choose for the decision making.

The above example shows the uncertainty in the output *F-N* curve that is caused by the uncertainty in the input parameters due to multiple teams. The uncertainty also exists, however, when the input parameters are elicited from one expert. It can also cause the uncertainty in the expectation of risk, which is to be studied in this research.

To make a more informative decision, we therefore must study the uncertainty in the input parameters and its impact on the uncertainty in the output [Morgan and Henrion 1990; Kumamoto and Henley 1996; Winkler 1996; Saltelli, Chan et al. 2000; Bedford and Cooke 2001; Abrahamsson 2002; Kurowicka and Cooke 2006]. It has has become a major concern to decision makers, especially those in the public arena [Morgan and Henrion 1990; Kurowicka and Cooke 2006]. Consequently intensive studies on uncertainty analysis have been seen in literature. Nuclear industry sees the earliest applications of PRA and the uncertainty analysis. A series of papers have been published related to the reactor safety [Helton, Johnson et al. 1995; Helton, Johnson et al. 1995; Helton, Johnson et al. 1995; Helton, Anderson et al. 1996; Helton, Bean et al. 1996; Helton, Bean et al. 1997; Helton 1999; Helton, Anderson et al. 2000; Helton, Martell et al. 2000; Kraan and Cooke 2000]. Other applications of the uncertainty analysis of PRA include chemical industry [Lauridsen, Kozine et al. 2002], offshore transport [Nilsen, Gudmestad et al. 1998], food safety [Frey and Patil 2001], and natural disaster analysis [Iman, Johnson et al. 2002; Bazzurro and Luco 2005; Li and Ellingwood 2006] etc. Uncertainty analysis is also used in other sectors such as cost analysis of engineering Project [Duffey and Van Dorp 1998] and production planning [Mula, Poler et al. 2006].

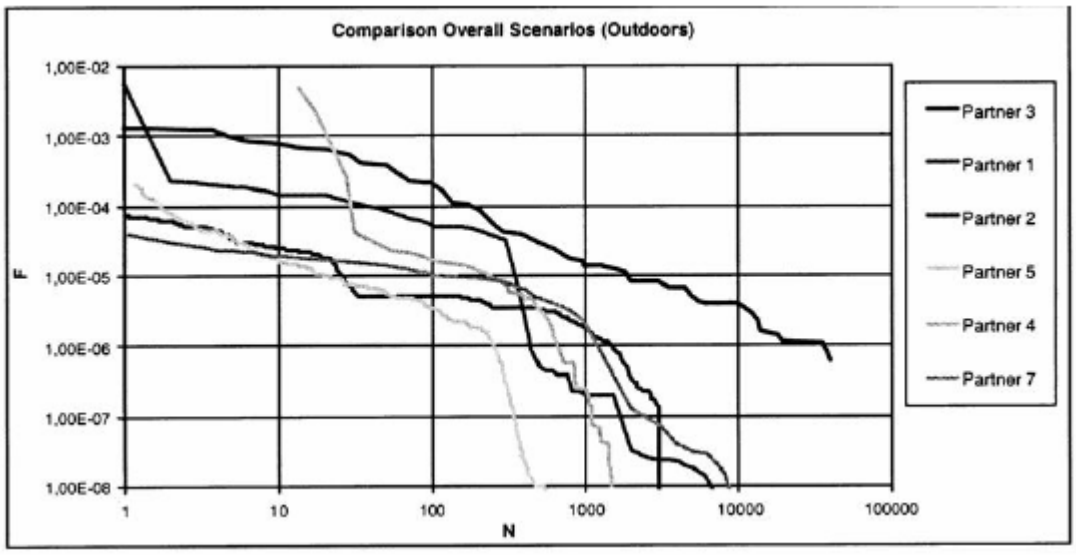


Fig. 1.6 Demonstration of the uncertainty in the estimated risk by F-N curves (reproduced with permission from [Abrahamsson 2002])

In addition to the uncertainty in the individual input parameters, dependence often exists among the input parameters. When two input parameters hold a strong positive dependency, they take either a high value or a low value simultaneously. Ignorance of such dependence can result in the significant underassessment of the uncertainty in the output. As a result, dependence among the input parameters has a strong impact on the uncertainty in the output. Therefore modelling the dependence among the input parameters holds an important role in uncertainty analysis [Duffey and Van Dorp 1998; Ferson and Hajagos 2005; Kurowicka and Cooke 2006]

With dependencies involved, we usually set up the uncertainty model of the input parameters in two steps [Kurowicka and Cooke 2006]. First, the uncertainty in each individual input parameter is usually assumed to follow a parametric distribution; the governing parameters of the distribution are then elicited through expert judgement. These parametric distributions are called the marginal distributions of the input parameters. Second, the dependence is modelled with the correlations or rank correlations that are elicited from experts as well. When the marginal distributions are assumed to be normal, the uncertainty model of the input parameters, i.e. a multivariate normal distribution, can be defined by further eliciting the correlation matrix. For other types of marginal distributions, the joint distribution cannot be defined by further eliciting the correlations or the rank correlations. In this case, the joint distribution is usually defined with the one of the minimum arbitrary information [Bedford and Cooke 2001]. An example of modelling the input parameters' uncertainty including the dependence can be found in [Helton, Johnson et al. 1995]. For the 34 input parameters in a reactor accident consequence model, each marginal distribution is assumed as either Uniform or Log-Uniform. Rank correlations between some pairs of input parameters are assessed for modelling the dependence. Another example can be found in [Helton, Anderson et al. 2000; Helton, Martell et al. 2000]. For the 57 input parameters in a PRA model WIPP (Waste Isolation Pilot Plant), each individual marginal distribution is assumed as either Uniform, or Log-Uniform, or Triangular, or Student's respectively. Rank correlations between some pairs of input parameters are assessed for modelling the dependence.

Modelling the dependence among the random variables is also an important issue in other application areas such as project risk analysis [Duffey and Van Dorp 1998; Van Dorp 2005], decision analysis [Clemen and Reilly 1999] and actuarial modelling [Frees and Valdez 1998].

1.5.3 General Procedure of Uncertainty Analysis

Suppose a mathematical model $Y = f(X)$, where X is a vector of the input parameters. Uncertainty analysis aims to assess the uncertainty in Y that is driven by the uncertainties in X [Morgan and Henrion 1990; Cooke 1997]. The uncertainty in X is usually represented by a joint probability distribution. The uncertainty in Y is then calculated by propagating the uncertainty in X through the model $Y = f(X)$. The general procedure for uncertainty analysis is shown in Fig. 1.7. In the example, the variable G is a function of three random variables with the distribution denoted as f_1, f_2 and f_3. Not shown in the diagram is the possible dependence among the three input parameters. The distribution of G is then derived through the model corresponding to the distributions of the input parameters. For a complex model, the distribution of the output is generally built by Monte Carlo simulations and therefore called simulated distribution. This procedure is also called probabilistic uncertainty analysis. With the strong foundations given by probability and statistics theories, probabilistic uncertainty analysis is by far the most widely used method [Abrahamsson 2002].

As a summary, the uncertainty analysis is composed of two steps. First, we build the uncertainty model in the input parameters. In the risk analysis context, it is usually done through the elicitation of expert judgement. Second, we calculate the uncertainty propagation through the model. Accordingly, we are going to identify our research aims for uncertainty analysis in the next section.

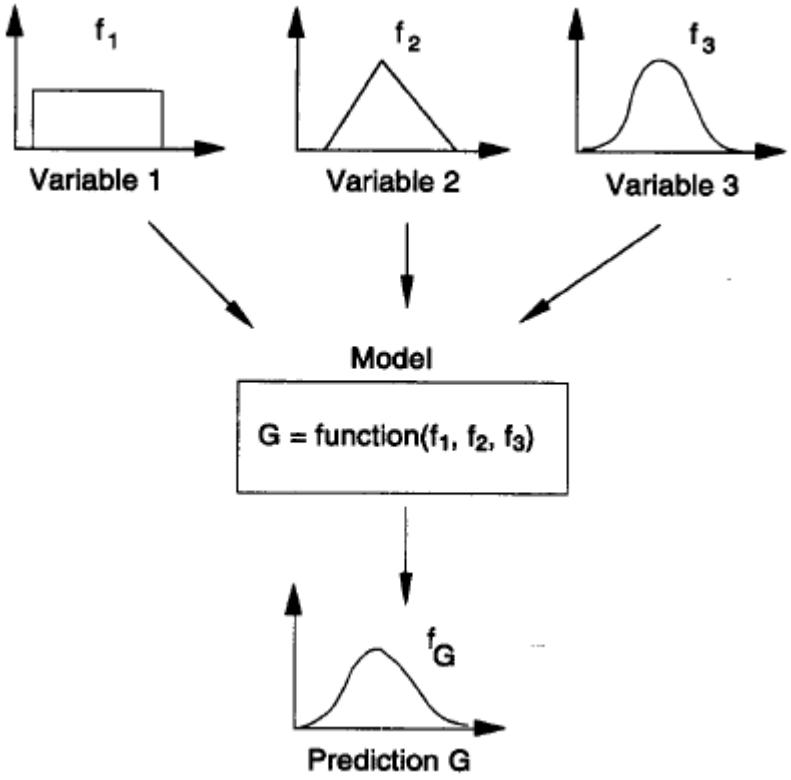


Fig. 1.7 Demonstration of uncertainty analysis of a mathematical model (reproduced with permission from [Abrahamsson 2002])

1.6 Research Aims

1.6.1 Overview

This research is founded by the Rail Safety and Standard Board (RSSB), UK. The goal is to develop methods for assessing the subjective uncertainty in the output of a PRA model. A PRA model usually is built up with three layers as demonstrated in Fig. 1.8. As we have discussed above, a PRA model is usually composed of fault trees and event trees. The fault trees and event trees are generally built up with computer software tools such as Isograph FT+. Therefore these fault trees and event trees are represented as computers codes that lie on the inner layer as shown in Fig. 1.8.

The fault trees and event trees are then parametralized usually through a combination of empirical data and expert judgement. Therefore the database of the parameters of

the fault trees and event trees form the parameter layer of the PRA model as shown in Fig. 1.8.

The experts usually make some common assumptions to assess the parameters. The narratives regarding the assumptions form the outside layer as shown in Fig. 1.8.

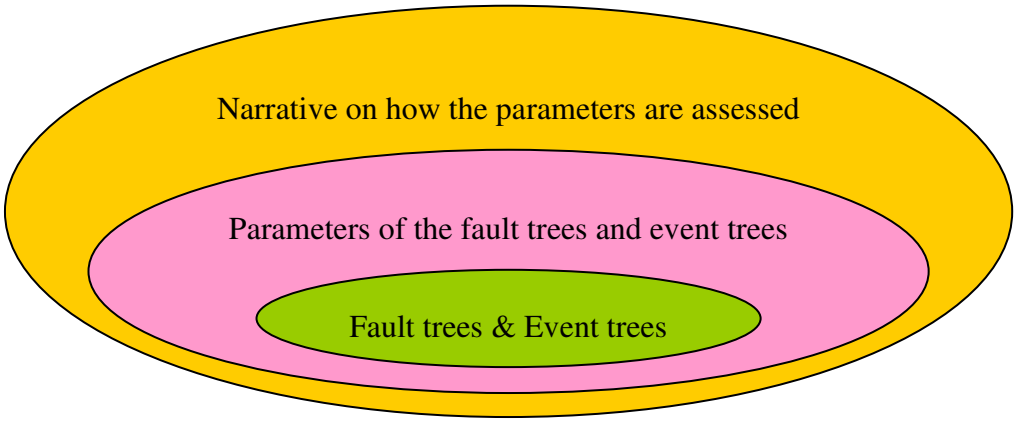


Fig. 1.8 Three layers of a PRA model.

We conduct the uncertainty analysis at the parameter layer. As discussed above, uncertainty analysis includes two steps: (1) build uncertainty models of the input parameters; and (2) compute the uncertainty propagation through the model. Generally there are three areas where input parameters are required in fault trees and event trees, including:

(1) the basic events or precursors of the fault trees;

(2) the consequences of the event trees; and

(3) the escalation events.

These three types of input parameters have different properties. At the first stage, we focus on studying the uncertainty in the basic events and the consequences; we treat as constants the probabilities of the escalation events.

Accordingly we identify three aims for this research:

(1) develop new methods for modelling the subjective uncertainty in the basic events;
(2) develop new methods for modelling the subjective uncertainty in the consequences;
(3) calculating the uncertainty in the output of the PRA, i.e. calculating the propagation of uncertainty of the input through the model.

We will discuss in details the above three research aims.

1.6.2 Modelling Subjective Uncertainty in Basic Events

There are many input parameters. States of knowledge uncertainties about these parameters are correlated. Consequently we have to model a high dimensional joint distribution that requires much elicitation time. It is essential to keep the elicitation time reasonable to the experts. We usually use parametric marginals for the uncertainty of the individual parameters and model the dependences among the input parameters separately. The most popular model is the multivariate normal distribution [Kurowicka and Cooke 2006]. Even assuming this model, there are still two issues however when we have many input parameters. Suppose we have n input parameters. Besides the means of all the input parameters, we need to elicit n standard deviations to define the marginals and $n(n-1)/2$ correlations to fill the correlation matrix. It is too much to do in practice when n is large. The second issue is that the outcome correlation matrix derived from elicitation must be positive semi-definite. It is a challenge if we fill in the correlation matrix cell by cell from elicitation [Bedford and Cooke 2001; Van Dorp 2005; Kurowicka and Cooke 2006].

Our aim is to develop a procedure for building the covariance matrix through expert judgement elicitation. The procedure should require reasonable time from experts. The procedure should be able to guarantee that the outcome covariance matrix is positive semi-definite.

1.6.3 Modelling Uncertainty in Consequences

The consequences of a risk scenario are defined here as the mean numbers of injuries, that is the product of the number of people exposed to the risk scenarios and the individual injury probability. In this research, the numbers of the people exposed to the risk scenarios are set as the point-estimates. These numbers are of course subject to uncertainty as well. For this research, however, we do not study the uncertainty in these numbers. Therefore we study the way that the uncertainty in the consequences is influenced by the uncertainty in the individual injury probabilities.

A study of typical fault tree models shows that individual injury probabilities are often decided by the outcomes of a subset of the escalation events. A combination of the outcomes of the escalation events in the subset that decides the individual injury probabilities is defined here as a rule set. Multiple risk scenarios can be associated with the same rule set. Once the individual injury probabilities are assessed for a rule set, they can be used for all the associated risk scenarios. Therefore modelling the consequences on the rule sets typically leads to a reduction of elicitation workload.

There are three difficulties in modelling the uncertainty in the rule sets. First, the rule sets are intensively interwoven to each other through the shared escalation events. It makes it very difficult to assess the correlation among rule sets. Second, the number of the rule sets increases exponentially with the number of the escalation events, which implies too many rule sets to assess separately for a large event tree. Third, the rule sets should satisfy the monotonicity property, i.e. a rule set associated with worse conditions should always have larger individual injury probabilities than another associated with better conditions. Such a monotonicity property cannot be modelled simply by defining the correlations among the rule sets.

In order to model uncertainty in the consequences, we must therefore develop suitable methods for modelling the uncertainty in the rule sets. The method should require reasonable elicitation time from the experts and be able to keep the monotonicity property among the rule sets.

1.6.4 Calculating Uncertainty Propagation through Fault-tree and Event-tree Models

Usually Fault-tree and Event-tree models are built with commercial software tools such as Isograph FT+. Consequently these fault trees and event trees are defined as the computer codes that are not very transparent to the users. For such a PRA model, it is difficult to manipulate the database underlying the model for calculating the uncertainty propagation. Furthermore, a large Fault-tree and Event-tree model can be composed of as many as thousands of input parameters. We need efficient methods for calculating the uncertainty propagation through such a large model.

Our aim is to select or develop efficient methods to conduct uncertainty propagation through a large PRA model. The methods should be able to work on the computer fault trees and events models built with commercial software tools. The methods should also be able to work efficiently on a large Fault-tree and Event-tree model.

1.7 Thesis Overview

The remainder of this thesis is organized as follows.

In Chapter 2, literature review is made on the following topics: (1) methods for eliciting the marginal probability distribution of the individual variables; (2) methods for eliciting the dependence between two input parameters; (3) architectures for modelling high dimensional dependence; and (4) methods for propagating the uncertainty through mathematical models. The gaps in the literature are then highlighted at the end of the chapter.

In Chapter 3, a procedure is developed for building the correlation matrix for multiple input parameters through elicitation on uncertainty factors. The input parameters are then partitioned into families by their definitions. A method is developed for deriving the variance of the input parameters within the associated families. Included at the end of the chapter are a method for eliciting the correlation

between two random variables and a method for eliciting the variance of an input parameter family.

In Chapter 4, a method is developed for modelling the uncertainty in the rule sets through injury atoms. An injury atom is defined for a hazard source. Once defined the injury atoms can be used for all the rule sets. The rule sets are then defined by the combination of the injury atoms.

In Chapter 5, a method is developed for building a mimic Excel model of the fault trees and event trees. With the mimic model, we can manipulate the data conveniently for conducting uncertainty analysis. We then select a linear transformation to define the correlated normal random variables on a set of independent standard normal random variables. The transformation is suitable when the covariance matrix is positive semi-definite and standard deviations are very small. An analytical solution of the variance is then developed. The software designs are then developed for calculating the analytical solution of the variance and for conducting the simulations based on the mimic Excel model.

In Chapter 6, two case studies are made with RSSB-SRM HET10 and HET12. The case studies validate the practical performance of the elicitation procedure and methods developed in this thesis. The case studies also validate that the methods for calculating the uncertainty propagation through a large Fault-tree and Event-tree model are efficient.

In Chapter 7, the aims of the research are reviewed first. The developments of this research are then summarized and discussed. At the end of the chapter some future research is proposed.

In Appendix A, we propose a set of methods for building the distribution of the sum of products of continuous random variables. More work however is needed to measure and control the approximation errors. Once this is done, these methods can

be used to calculate the analytical approximate distribution of output of a Fault-tree and Event-tree model.

In Appendix B, tables giving complete information relating to the cases considered in Chapter 6 are presented.

In Appendix C, some theories on positive definite matrix and positive semi-definite matrix are presented. These theories are referred to in Chapter 3.

Chapter 2

Literature Review

2.1 Introduction

As stated in Chapter 1, this research focuses on the subjective uncertainty analysis of a mathematical model, which mainly involves two tasks: (1) modelling the subjective uncertainty in the input parameters through expert judgement elicitation; and (2) propagating uncertainty through the mathematical model.

In the literature, expert judgement is often used for building the distributions for individual variables, i.e. the marginal distributions. Due to the difficulties of assessment, dependence was initially ignored in building the uncertainty of multiple input parameters [Smith, Ryan et al. 1992]. It was found, however, that dependence had a strong effect on the probability distribution of the output and therefore must be modelled for most applications [Clemen and Winkler 1985]. By the divide-and-conquer strategy, the elicitation of dependence can be separated from the elicitation of the marginal probability distributions of the individual variables [Clemen, Fischer et al. 2000]. Theoretically the dependence can be modelled by the elicitation of conditional distributions. The elicitation of the conditional distribution is however difficult even for two variables. It is practically impossible to elicit the conditional distributions for a set of variables [Ravinder, Klenmuntz et al. 1988]. A feasible way is to model the dependence between two variables with the correlation coefficient or rank correlations. For multiple variables, the dependence between each pair of variables can be elicited. The pairwise dependencies are then organized to define the dependence among the multiple variables. The structure used to organize the pairwise dependencies is called dependence structure [Clemen, Fischer et al. 2000;

Kurowicka and Cooke 2006]. The correlation matrix, for example, is a common dependence structure.

As shown in Fig. 2.1, the literature review is made to cover four parts: (1) the methods for eliciting the marginal probability distribution of the individual variables; (2) the methods for eliciting the dependence between two input parameters; (3) the dependence structures for modelling high dimensional dependence; and (4) the methods for propagating the uncertainty through mathematical models. In this literature review we compare the existing methods for elicitation from a practical perspective. The aim is to select suitable elicitation methods and/or to find where we can put our efforts to improve the methods for the use in our project. Rather than investigating the psychological and behavioural theories underlying the elicitation methods, we compare the elicitation methods based on the current knowledge of behavioural theories and the applications in elicitation practices so far. At the end of the chapter, we highlight the gaps that we find in the literature.

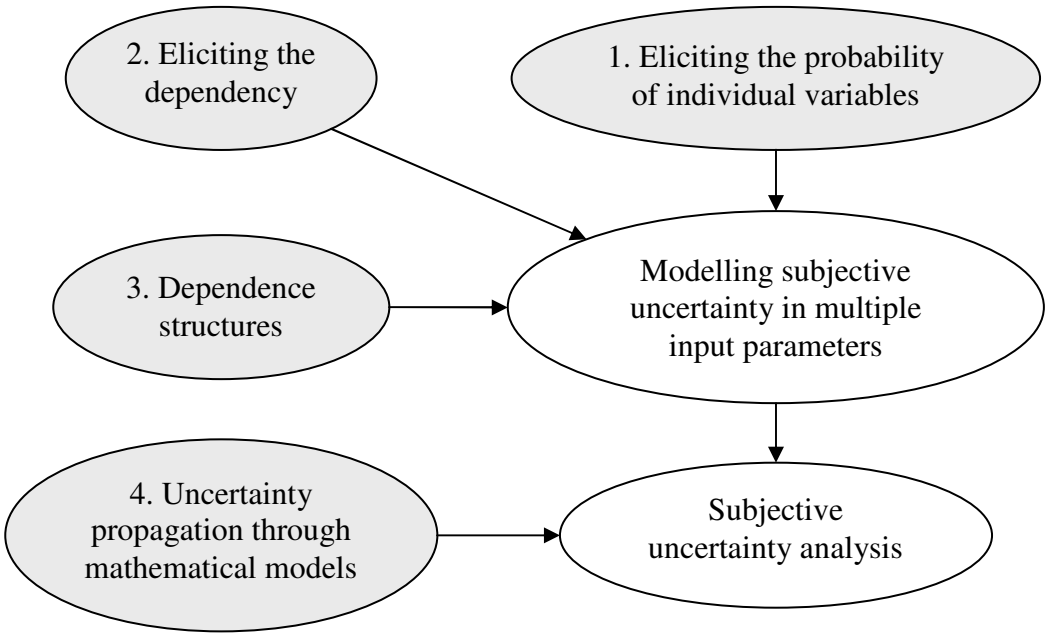


Fig. 2.1 Demonstration of the literature review scope

2.2 Modelling Subjective Uncertainty in Multiple Input Parameters

To assess the uncertainty in the output of a mathematical model, we need to build a uncertainty model for the input parameters [Morgan and Henrion 1990]. In the risk analysis context, we usually build the subjective uncertainty in the input parameters through expert judgement. The subjective uncertainty represents the experts' knowledge and beliefs about the uncertainty in the input parameters. Elicitation is the process by which we formulate the experts' knowledge and beliefs into probability distributions [Garthwaite, Kadane et al. 2005]. It is admitted that there is a limit in experts' ability to perceive abstract variables. With the great effort in improving the elicitation, however, expert judgement is increasingly employed for input in quantitative uncertainty modelling [Kurowicka and Cooke 2006; O'Hagan, Buck et al. 2006].

A sound process is essential to implement a good elicitation. The elicitation process proposed by SRI International for eliciting the probabilities [Merkhofer 1987] is composed of seven stages. In line with the SRI International process, a new process is developed to cope with bias caused in the elicitation [Walls and Quigley 2001]. The process includes five main stages. Another process with five stages is summarized via the transcript of a real elicitation interview [Shephard and Kirkwood 1994]. Another elicitation process with seven steps is recommended in [Clemen and Reilly 2001]. Broadly these proposed processes cover the same tasks. Some tasks may be split into two in some processes; the order in which the tasks should be carried out may be slightly different. At a high level, four basic stages for implementing an elicitation are summarized as: *Set up, Elicit, Fit*, and *Evaluate* [Garthwaite, Kadane et al. 2005]. The *Set up* stage consists of the preparation for the elicitation. Specifically it includes: selecting the experts, training the experts, identifying what aspects of the problem to elicit. The *Elicit* stage is the implementation of elicitation. The experts are asked to answer the properly designed questions. In the *Fit* stage the elicited information, i.e. the experts' answers, is transferred into the mathematical models such as the distributions or joint distributions. In the *Evaluate* stage, we need to judge whether the elicitation outcome

is adequate or not. We need to emphasize that an elicitation is well done if the elicited information accurately represents the expert's belief and knowledge. It is not related to how good that knowledge is [Cooke 1991; Garthwaite, Kadane et al. 2005].

From the description of the elicitation process, one important issue about the elicitation is what kinds of variables can be elicited from the experts. As a guiding principle, experts should be asked the questions about quantities that are meaningful to them. This suggests that questions should generally concern observable quantities rather than unobservable parameters [Garthwaite, Kadane et al. 2005; O'Hagan, Buck et al. 2006]. In some applications, it is also possible that the particular statistical model is so familiar to the experts that their parameters have acquired well-understood scientific meaning. In this case it is also appropriate to ask experts directly about such parameters [Kadane 1980; Winkler 1980]. For instance, the probability of the basic event of a fault tree has a well understood meaning and therefore is not an abstract model parameter. Therefore it can be elicited from the experts [Bedford and Cooke 2001].

2.3 Eliciting Marginal Distributions

2.3.1 Overview

The uncertainty about an unknown continuous variable is generally expressed by its probability distribution, called marginal distribution. The marginal distribution is usually inferred from the elicitation of some summary statistics of the unknown variable [Garthwaite, Kadane et al. 2005; O'Hagan, Buck et al. 2006]. The summary statistics widely found in the literature include probabilities, quantiles, intervals, the modes and the medians [Garthwaite, Kadane et al. 2005; O'Hagan, Buck et al. 2006]. Details on these summaries will be discussed later.

In practice we can only elicit a few summary statistics from the experts. The marginal distribution can then be built by fitting the elicited summary statistics based on some assumptions. For example, we can assume that that the cumulative distribution function (CDF) is continuous and smooth. Consequently a nonparametric

marginal distribution can be fitted with the elicited summary statistics. A good summary of the methods for doing such fitting can be found in [O'Hagan, Buck et al. 2006]. To get a reasonably good representation of the distribution, quite a few summary statistics are required. The marginal distribution can also be assumed to follow a parametric distribution family. It mainly has three advantages: (1) fewer assessments of the summaries are required; (2) expert judgement can be checked and adjusted by a few extra assessments; (3) parametric distributions are easy to analyze.

2.3.2 Summary Statistics

The cumulative probabilities of an unknown variable are valuable summaries. Suppose that X is an unknown variable. We can set a series of values of $X = x_i$, $i = 1, \cdots, n$ and $x_i < x_j$ for $i < j$. The probabilities $\Pr(X \le x_i)$ are then elicited from the experts. Alternatively, we can set a series of probabilities spreading over the interval $[0, 1]$ that are denoted as p_i, $i = 1, \cdots, n$. The values $x(p_i)$ satisfying $\Pr(X \le x(p_i)) = p_i$, called quantiles or percentiles, are then elicited from the experts. The most used quantile is the median, which is obtained by setting $p_i = 0.5$. Starting with the median, a method of bisection is often used to elicit a number of quantiles [O'Hagan, Buck et al. 2006]. By the bisection method, the expert is firstly asked the question:

Q1. Can you determine a value such that it is equally likely that X is less than or greater than this value?

The elicited value from this question is the median, i.e. $x(0.5)$. We will then ask the expert the two more questions:

Q2. Suppose that X is below $x(0.5)$. Can you now determine a new value such that it is equally likely that X is less than or greater than this value?

Q3. Suppose that X is above $x(0.5)$. Can you now determine a new value such that it is equally likely that X is less than or greater than this value?

The answers to the above two questions give the experts' assessments of $x(0.25)$ and $x(0.75)$. This can continue with more bisections, although it becomes more difficult for the expert to assess. A good example of the bisection method can be found in [Peterson, Snapper et al. 1972].

The quantile elicitation can also be carried out on the so-called credible interval [Garthwaite, Kadane et al. 2005; O'Hagan, Buck et al. 2006]. Associated with a probability $p \in [0,1]$, the central credible interval is defined by $[x(0.5-p/2), x(0.5+p/2)]$, where $x(0.5-p/2)$ and $x(0.5+p/2)$ are the two quantiles that satisfy

$$\Pr(X \le x(0.5-p/2)) = 0.5-p/2$$

and

$$\Pr(X \le x(0.5+p/2)) = 0.5+p/2$$

For example, the central credible interval associated with $p = 0.5$ is the interval defined by the lower and upper quartiles.

In elicitation, the expert is advised with the definition of the central credible interval and then is asked to assess $x(0.5-p/2)$ and $x(0.5+p/2)$ corresponding to the probability p. Many experiments have demonstrated that the experts perform reasonably well in assessing credible intervals. However there is a clear tendency for experts to give a short central interval, which is called overconfidence [Peterson, Snapper et al. 1972; Schaefer and Borcherding 1973; Lichtenstein and Fischhoff 1980].

So far there is still no conclusion on whether eliciting cumulative probability or eliciting quantiles leads to better performance. Eliciting the probability is reported as giving better performance in [Seaver, Von Winterfeldt et al. 1978], while eliciting the percentiles is reported as giving better performance in [Murphy and Winkler 1974]. More recently an adaptive fixed interval method is reported as giving better performance in [Winman, Hansson et al. 2004], while eliciting the percentile including the median is recommended in [Soll and Klayman 2004].

As an alternative to eliciting the probabilities and the quantiles, the ratios of the probabilities can be elicited in the method termed the analytic hierarchy process (AHP)[Saaty 1977; Saaty 1980; Hughes 1993; Cagno, Caron et al. 2000; Monti and Carenini 2000]. To elicit a set of probabilities $p_1, p_2, \cdots, p_n$, the expert is asked to assess the relative probabilities p_i / p_i for $i \neq j$. The probabilities p_i, $i = 1, \cdots, n$ can be derived by eliciting the minimum set of $(n-1)$ ratios, e.g. p_i / p_1, $i = 2, \cdots, n$. When more ratios are elicited, the inconsistency among the assessments can be identified and used to adjust the expert judgment [Saaty 1980; Basak 1998]. Rather than eliciting the absolute value, the ratios are assessed using a scale from 1 to 9, where the points on the scale are associated with verbal descriptions. For example, the ratio $\Pr(A)/\Pr(B) = 5$ is described as "A is strongly more probable than B". Therefore we can see that AHP is based on relative verbal assessments that are believed easy to implement in practice [Clemen, Fischer et al. 2000; Garthwaite, Kadane et al. 2005; O'Hagan, Buck et al. 2006; Park and Lee 2008]. The AHP methods are still under development with the enhancement by other techniques. A Bayesian paired comparison approach is developed for assessing the accident probability in [Szwed, Van Dorp et al. 2006]. A new method for eliciting human error probabilities is developed as a combination of AHP and the success likelihood index method (SLIM) [Park and Lee 2008].

Psychological scaling models utilize paired comparisons to implement elicitation as AHP does [Cooke 1991]. A set of events are pairwise compared in terms of their likelihood. The elicited data are then analyzed with a psychological model. Based on different assumptions, three psychological scaling models are developed including:

the Thurstone model, the Bradley-Terry model, and the NFL(negative exponential lifetime) model [Cooke 1991]. Compared with AHP, the Psychological scaling models use clear statistical assumptions.

Besides the summaries related to probability, an expert's ability to estimate the location measures including the mean, the median and the mode has been investigated in some experiments [Spencer 1961; Spencer 1963; Peterson and Miller 1964; Beach and Swenson 1966]. The experiments show that the expert can give good estimation for all three location estimators for an approximately symmetric distribution [Spencer 1961; Beach and Swenson 1966]. However for highly skewed variable, the experiment shows that the expert has good performance in assessing the median and the mode while the mean is biased towards the median [Peterson and Miller 1964].

2.4 Methods for Eliciting Dependence between Two Random Variables

2.4.1 Overview

As presented before, the dependence between two variables is modelled practically by eliciting the correlation coefficient. In this section, we start with a brief description of the three correlation coefficients that are commonly used to measure the dependence between two variables. As for eliciting the marginal distributions of the individual variables, the correlation coefficients need to be derived by eliciting some summary statistics regarding the dependence. The correlation itself is one of the summaries that can be elicited. Six elicitation methods found in the literature are reviewed.

2.4.2 Classification of Correlations

The three types of correlations defined for two random variables are product-moment correlation, rank correlation and Kendall's τ. Product-moment correlation is also called Pearson's correlation. For two random variables X and Y, with finite

expectations $E(X)$, $E(Y)$ and finite variance σ_X^2 , σ_Y^2 , the product-moment correlation is defined as:

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

Product-moment correlation defined above is a measure of the linear relationship between the random variables and $r(X,Y)=\pm 1$ if and only if $Y=aX+b$ for some $a\neq 0$.

Rank correlation is also called Spearman rank correlation. Suppose that *X* and *Y* are two random variables. Suppose further that F_X and F_Y are the cumulative distribution functions of *X* and *Y* respectively. Define the quantile variables

$$U=F_X(X)$$

$$V=F_Y(Y)$$

By the above definition, *U* and *V* are two random variables with the support defined on $[0,1]$. The rank correlation of *X* and *Y* is then defined as

$$\rho_r(X,Y)=r(U,V)=\frac{E(UV)-E(U)E(V)}{\sigma_U\sigma_V}$$

Therefore rank correlation is defined on the quantile variables. As a result, rank correlation is invariant with respect to strictly increasing transformations of the random variables, as all such transformations have the same quantile variables.

Suppose that *X* and *Y* are two random variables of the cumulative distribution function F_X and F_Y respectively. Let (X_1,Y_1) and (X_2,Y_2) be two independent pairs of samples of (X,Y). Kendall's τ of *X* and *Y* is defined as

$$\tau = \Pr[(X_2 - X_1)(Y_2 - Y_1) > 0] - \Pr[(X_2 - X_1)(Y_2 - Y_1) < 0]$$

For two random variables *X* and *Y*, the above three correlations can be derived through the elicitation methods discussed below. When *X* and *Y* are assumed to follow normal distributions, the joint normal distribution can then be defined with the product-moment correlation. For the general marginal distributions, the joint distribution of *X* and *Y* can be derived with a suitable copula that is to be discussed later on.

2.4.3 Main Methods for Eliciting Correlations

Six main methods for eliciting the correlations have been well investigated in the literature. As summarized in [Clemen, Fischer et al. 2000] the six methods include:

- the correlation strength, called the *S* method;
- direct elicitation of the correlation coefficient, called the *R* method;
- conditional fractile elicitation, called the *CF* method;
- concordance probability elicitation; called the *CNC* method;
- joint probability elicitation, called the *JP* method;
- conditional probability elicitation, called the *CP* method

With the *S* method, a continuous line scale is presented; and the left and right ends represent being independent and perfect correlation respectively. The expert is asked to mark on the line with the location representing his belief of the dependence between the two variables. The location of the mark is then linearly transformed to a correlation between 0 and 1. In addition, the expert is asked whether the two variables are positively or negatively correlated.

With the *R* method, the expert is asked to assess directly the correlation coefficient between the two random variables.

Suppose the random variables X and Y have marginal distributions $F(x)$ and $G(y)$ respectively and the joint density $f_{X,Y}(x,y)$. With the *CF* method, the expert is informed that a pair of sample (x,y) is drawn randomly from $f_{X,Y}(x,y)$. The expert is then informed about the probability $\Pr(Y \le y)$, i.e. $G(y)$. In the case studies in [Clemen, Fischer et al. 2000], it is set as $G(y)=0.9$. The expert is then asked to assess probability $\Pr(X \le x | Y \le y)$. The correlation between X and Y can then be derived through non-parametric regression:

$$E\left(F_{X|Y\le y}\right) = r(X,Y)\left(G(y)-0.5\right)+0.5$$

In the above formula, $G(y)$ is preset; $E\left(F_{X|Y\le y}\right)$ is obtained from the elicitation; and the Spearman's rank correlation $r(X,Y)$ can then be solved in terms of $G(y)$ and $E\left(F_{X|Y\le y}\right)$. We can see that $G(y)$ has to be taken not equal to 0.5.

With the *CNC* method, the expert is informed that two pairs of samples (x_1,y_1) and (x_2,y_2) are drawn randomly from $f_{X,Y}(x,y)$. The expert is then informed that $x_2 > x_1$ and is asked to assess the probability of $y_2 > y_1$. It is the concordance probability that is actually elicited by the *CNC* method. Naturally Kendall's τ is employed to model the elicited probability as defined as

$$\tau = 2P_C - 1$$

where the P_C stands for the elicited probability; τ is the estimate of Kendall's τ.

The derived Kendall's τ can be used directly to define the joint distribution $f_{X,Y}(x,y)$ by selecting a suitable copula [Bedford and Cooke 2001; Kurowicka and Cooke 2006]. When the bivariate normal distribution is assumed for X and Y, the

Pearson product-moment correlation and the Spearman's rank correlation can also be derived from the Kendall's τ as defined in [Kruskal 1958] as

$$r(X,Y) = \sin\left(\pi\tau/2\right)$$

$$r(X,Y) = 2\sin\left(\pi\rho_r/6\right)$$

where $r(X,Y)$ and ρ_r stand for the Pearson's correlation and Spearman's rank correlation respectively.

With the *JP* method, the expert is asked to assess the probability of a random sample falling in a specific area, i.e. the probability $\Pr(X \le x_L \text{ and } Y \le y_L)$, where x_L and y_L are preset values such as the 30 percentile [Clemen, Fischer et al. 2000].

With the *CP* method, the expert is told that sample (x, y) is drawn randomly from $f_{X,Y}(x, y)$. The expert is then informed *x* falls into the lower $P \in [0,1]$ part of *X*, i.e. $x \le x_P$, where x_P is the $100 \times P$ percentile of *X*. The *P* is set as 60% in the case studies in [Clemen, Fischer et al. 2000]. The expert is then asked to assess the probability $\Pr(Y \le y_P | X \le x_P)$, where y_P is the $100 \times P$ percentile of *Y*. With the elicited joint probability from the *JP* method or the conditional probability from the *CP* method, the Pearson correlation between *X* and *Y* can then be calculated by assuming bivariate normal distribution for *X* and *Y*.

The above six methods are compared with two case studies in [Clemen, Fischer et al. 2000]. The research demonstrates that the *R* method and the *S* method give the best performance in terms of accuracy. They are also considered the easiest to implement. This conclusion seems controversial since the *R* method was reported as a poor one in the earlier research [Gokhale and Press 1982; Morgan and Henrion 1990; Kadane and Wolfson 1998]. For the *S* method, it seems a strong assumption to linearly transform the mark location to the correlation. The good performance of the *R*

method and the *S* method in [Clemen, Fischer et al. 2000], however, can be attributed to the two aspects: (1) the respondents in the experiments are very familiar with the statistical correlations; (2) the training method is very efficient. The training was done in the same way as the software *Crystal Ball* does [ORACLE 2008]. For a pair of bivariate normal random variables, the samples are drawn on a preset correlation and then the scatter plot is presented on the screen. Based on the scatter plot, the expert is asked to assess the correlation by the *S* method and *R* method. The expert is then informed with the true correlation to adjust their judgment. The investigation in [Clemen, Fischer et al. 2000] demonstrates that the good assessment of the correlations can be obtained by the *S* and *R* method when the expert knows well about the statistical correlations and has an efficient training process.

2.5 High Dimensional Dependence Modelling

2.5.1 Overview

Considering the difficulties and the cost in elicitation, the uncertainty model for multiple input parameters in practice is usually built by parametric marginal distributions of the individual variables plus the correlations among the parameters. Consequently, the uncertainty model can be built in two stages: (1) building the marginal distribution of the individual parameters; (2) modelling the dependences among the parameters [Kurowicka and Cooke 2006]. After the review of the methods for eliciting the marginal probability distributions and the correlations between a pair of random variables, the methods for defining the correlations for high dimensional input parameters are to be reviewed in this section.

2.5.2 Multivariate Normal Distributions

When the individual input parameters can be assumed to follow normal distributions, the uncertainty in the input parameters can be completely defined by the means, the variances, and the correlation matrix. The multivariate normal distribution is one of the best choices for modelling the uncertainty in the input parameters. It requires less

elicitation than other parametric joint distributions do. It is also easy to conduct analysis [Ghosh and Henderson 2003; O'Hagan, Buck et al. 2006].

For n random variables, there are $n(n-1)/2$ correlation coefficients needed to elicit to build the correlation matrix. For very high dimensional input parameters, however, elicitation work for building the correlation matrix is still too time intensive. The situation becomes worse since we have to keep the outcome matrix positive definite or positive semi-definite. Different methods have been developed to obtain a positive semi-definite correlation matrix by adjusting the outcome matrix from elicitation. Usually some assumptions have to be made for this purpose [Ghosh and Henderson 2002; Ghosh and Henderson 2003; Kurowicka and Cooke 2006].

2.5.3 Copula Trees and the Copula Vines

For two random variables with general marginal distributions rather than normal distributions, the joint distribution cannot be defined completely by defining the correlation matrix. This difficulty can be dealt with by introducing copulas that are defined on the quantile variables.

Suppose that the random variables X and Y have cumulative distribution functions $F_X(x)$ and $F_Y(y)$ respectively. Define $U = F_X(X)$ and $V = F_Y(Y)$. Therefore U and V are random variables of uniform marginal distributions. The copula of X and Y is defined as the joint distribution of (U, V) and denoted as $C_{X,Y}(u,v)$. Every continuous bivariate distribution can be completely determined by the copula and the marginal distributions as $F_{X,Y}(x,y) = C_{X,Y}(F_X(x), F_Y(y))$ [Clemen and Reilly 1999]. For a given copula $C_{X,Y}(u,v)$, Spearman's rank correlation coefficient and Kendall's τ can be calculated as in [Nelsen 2006]. We know that the transformation by the cumulative distribution function does not change Spearman's rank correlation or Kendall's τ. Therefore we can design the copula $C_{X,Y}(u,v)$ to satisfy the rank correlations that are elicited for X and Y. As a result, the outcome joint distribution that is defined by the $F_X(x)$, $F_Y(y)$ and $C_{X,Y}(u,v)$ has the rank correlation as

elicited from the expert [Jouini and Clemen 1996; Clemen and Reilly 1999; Bedford and Cooke 2001; Kurowicka and Cooke 2006; O'Hagan, Buck et al. 2006]. Case studies using copulas to define dependence can be found in [Clemen and Reilly 1999].

Based on the specific characteristics, different copulas can be defined. A good general review of the copulas can be found in [Bedford and Cooke 2001; Kurowicka and Cooke 2006; Nelsen 2006].

For high-dimensional input parameters, Markov trees and copula vines can be used to define the joint distributions on the rank correlations or conditional rank correlations [Bedford and Cooke 2001; Bedford and Cooke 2002]. The outcome joint distribution can satisfy exactly the rank correlations elicited from the experts. Many case studies using Markov trees and copula vines have been found in the literature such as [Yi and Bier 1998; Clemen and Reilly 1999; Kurowicka and Cooke 2006].

Markov trees cannot be used to represent any joint dependence structure. Copula vines however can be constructed to model any given rank correlation matrix exactly.

2.5.4 Modelling Dependence through Common Factors

The method for modelling the dependence through common factors was initially developed on a single common factor in a risk analysis model in [Duffey and Van Dorp 1998]. It was further extended to multiple factors in [Van Dorp 2005]. As demonstrated in Fig 2.2, the U_i's in the left side circles represent common factors to the input parameters X_j's in the right side circles; the variable Y_j in the middle box is a surrogate variable aggregating the effects of the U_i's on the X_j's.

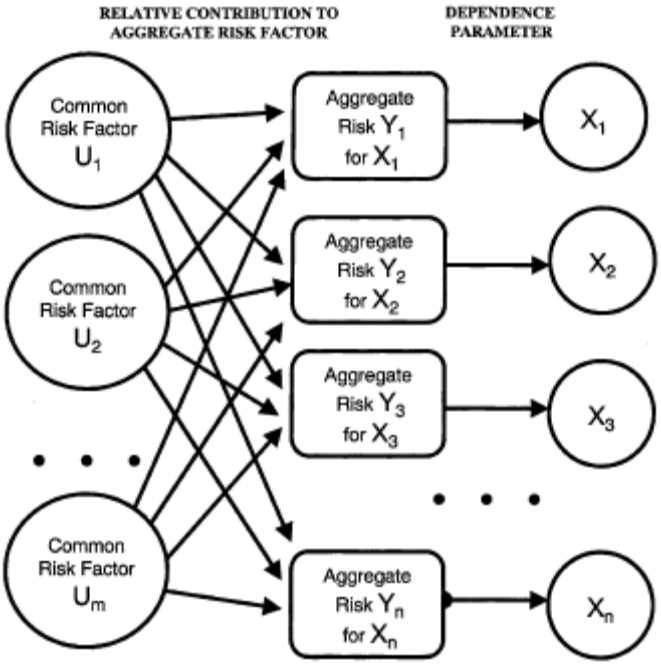


Fig. 2.2 A model for defining the dependence among multiple random variables on the common factors, extracted from [Van Dorp 2005]

The common factors U_i are assumed to be uniformly distributed on the interval $[0,1]$. The lowest and highest risk level of U_i are set to 0 and 1 respectively. To one input parameter X_j, the factors U_i, $i = 1, \cdots, m$ can have different importance. To measure the relative importance of the factors to X_j, the weights $w_{j,i}$, $i = 1, \cdots, m$, are defined. To derive the weights, the *m* factors can be compared pairwise in terms of their importance to the input parameter X_j. For each pair of factors $U_{i(1)}$ and $U_{i(2)}$, $i(1), i(1) = 1, \cdots, m$, the elicited information is how many times $U_{i(1)}$ is more important to X_j than $U_{i(1)}$. With the elicited information, the weights can be derived and normalized to keep

$$\sum_{i=1}^{m} w_{j,i} = 1,\ 0 \le w_{j,i} \le 1$$

Therefore the weights measure the relative importance of a factor to one input parameter. We can say that $U_{i(1)}$ is $\frac{w_{j,i(1)}}{w_{j,i(2)}}$ times more important to X_j than $U_{i(2)}$.

Based on the elicited weights, the aggregated factor is defined as a weighted linear combination

$$Y_j = \sum_{i=1}^{m} w_{j,i} U_i$$

At the next step, the bivariate dependence between Y_j and X_j needs to modelled. It can be implemented with a copula $C\left(F\left(Y_j\right), F\left(X_j\right)\right)$.

As a summary, the dependence between X_j, $j = 1, \cdots, n$ is modelled by the common factors U_i, $i = 1, \cdots, m$, the weights, the aggregated factor Y_j's and the copulas $C\left(F\left(Y_j\right), F\left(X_j\right)\right)$. Modelling through common factors offers a flexible framework that does not require a special structure among the parameters. Exploring and ranking the common factors U_i's offer the insight into the dependence mechanism in a structured way. To populate the model, $n \times m$ weights $w_{j,i}$'s and n copulas $C\left(F\left(Y_j\right), F\left(X_j\right)\right)$ need to be elicited. For high-dimension models, it requires much less elicitation workload than directly eliciting the $n(n-1)/2$ correlations among the n parameters. Since no correlation matrix of the parameters is required, there is no difficult in keeping it positive definite.

In this method, if the factors U_i were independent normal random variables, the aggregated factors Y_j would be correlated normal random variables. From this perspective, the method is close to the idea of a normal copula.

2.6 Uncertainty Analysis Methods

2.6.1 Overview

There are two types of uncertainties: model uncertainty and parameter uncertainty [Morgan and Henrion 1990]. Model uncertainty arises from the fact that any model is a simplification of the reality. It is not covered in this research. We focus on parameter uncertainty.

Suppose there is a mathematical model $Y = f(\mathbf{X})$, where Y is a scalar output; $\mathbf{X} = [X_1, \cdots, X_n]$ is a vector of the input parameters of the dimension n. Parameter uncertainty analysis aims to assess the uncertainty in y that is caused by the uncertainties in $\mathbf{X}$ [Rabitz 1989; Cooke 1997; Saltelli, Chan et al. 2000]. It is also called uncertainty propagation in [Morgan and Henrion 1990; Iman, Johnson et al. 2002]. The uncertainty in $\mathbf{X}$ is represented by a joint probability distribution. The uncertainty in Y is then calculated as the propagation of the input uncertainty through the model, and called probabilistic uncertainty analysis [Abrahamsson 2002; Helton, Johnson et al. 2006]. There are also some other types of uncertainty representations such as interval analysis and fuzzy theory [Helton, Johnson et al. 2004]. We focus on probabilistic uncertainty analysis in this research.

In this section, a literature review is made on the methods for probabilistic uncertainty analysis, including analytical methods, and simulation-based methods.

2.6.2 Analytical methods

With the joint distribution of the input parameters, the distribution of the output should ideally be derived as the distribution of the function of the random variables [Ross 2003]. This is called the analytical method. The exact analytical distribution of the output however can be derived only for simple models such as linear models of normal variables. Generally the original model needs to be approximated with its Taylor series. The mean, the variance and the higher order moments can then be calculated based on the approximate model. Since only the moments of the output

can be derived, the method is also referred to as the "method of moments" [Morgan and Henrion 1990; Robinson 1998; Abrahamsson 2002].

We can see that the "method of moments" only represents an approximation of the original model. When the variance in the input is large, higher order terms in the Taylor expansion must be included. This introduces much more complexity of the algebra especially for the complex original models, as is the case in risk analysis.

2.6.3 Simulations-based Methods

2.6.3.1 Overview

Given the limitation of the analytical methods, simulation-based methods are the most popular ways to carry out probabilistic uncertainty analysis. This trend has been enhanced further by the rapid development of the computer and software technologies.

Suppose that the joint distribution of $\mathbf{X} = [X_1, \cdots, X_n]$ is known. A sample of $\mathbf{X}$ is denoted as $\mathbf{x}_k = [x_{k,1}, \cdots, x_{k,n}]$, where $x_{k,i}$ is a sample of X_i. By the joint distribution of $\mathbf{X}$, we can generate a set of independent and identically distributed samples of $\mathbf{X}$, denoted as $\mathbf{x}_k$, $k = 1, \cdots, n_s$, where n_s stands for the size of the samples. The sample $\mathbf{x}_k$, can then be fed into the model $f(\mathbf{X})$ and a sample of Y can be calculated as $y_k = f(\mathbf{x}_k)$. Such a process is called Monte Carlo simulation. Based on the simulated y_k, $k = 1, \cdots, n_s$, the mean, the variance and the empirical distribution of Y can then be estimated. This is called a simulation-based method. [Helton and Davis 2002]

From the above description, the core of the simulation-based method is to generate samples of the input parameters that satisfy the pre-specified joint distribution. The two most important sampling methods are random sampling and Latin hyper-cube sampling [Helton and Davis 2000; Helton, Johnson et al. 2006].

2.6.3.2 Random Sampling

Random sampling is associated with Monte-Carlo simulation [Sobol 1974]. Suppose a random variable X has the CDF $F_X(x)$. By the random sampling method, the samples of X is generated using $F_X(x)$. Define the random variable U as uniformly distributed in [0, 1]. At the first step, a set of independent and identically distributed samples of U, denoted as $u_1,\cdots,u_{n_s}$, are generated through a algorithm, called random number generator. The most common algorithm is the multiplicative congruential method. Starting with a seed s_0, a sequence of numbers are generated by the multiplicative congruential method as

$$s_k = (as_{k-1}) \bmod (m),\ k = 1,2,\cdots$$

where a and m are preset integers; mod( ) is the function returning the remainder after as_{k-1} is divided by m.

Therefore s_k takes a value in $\{0,\cdots,m-1\}$. The samples of U are then calculated by

$$u_i = \frac{s_k}{m},\ k = 1,\cdots,n_s$$

With large a and m, $u_1,\cdots,u_{n_s}$ are approximately independent and identically distributed samples of U. The IBM System/360 Uniform Random Generator is built on the multiplicative congruential method taking $m = 2^{31} - 1$ and $a = 7^5$. It has been used in many mathematical software packages and gives good results.

We can see that the $u_1,\cdots,u_{n_s}$ are not really random; they are decided by the preset seed s_0 and the preset integers a and m. Therefore $u_1,\cdots,u_{n_s}$ are called pseudo-random numbers. More discuss on generating the pseudo-random numbers can be found in [L'Ecuyer 1998].

At the second step, the samples of X can be generated based on $u_1, \cdots, u_{n_s}$ as

$$x_k = F_X^{-1}(u_k),\ k = 1, \cdots, n_s$$

where

$$F_X^{-1}(u) = \inf\{x : F_X(x) \geq u\}$$

is the inverse function of $F_X(x)$.

When the multiple input parameters $\mathbf{X} = [X_1, \cdots, X_n]$ follow a joint normal distribution, the sampling can be done through the *Cholesky* decomposition [Scheuer and Stoller 1962; Law and McComas 1999]. Suppose

$$E(\mathbf{X}) = E(X_1, \cdots, X_n)^T = (\mu_1, \cdots, \mu_n)^T$$

$$\mathrm{cov}(\mathbf{X}) = \mathrm{cov}(X_1, \cdots, X_n) = \Sigma_{\mathbf{x}}$$

When $\Sigma_{\mathbf{x}}$ is positive semi-definite, it can be factored by the *Cholesky* decomposition as

$$\Sigma_{\mathbf{x}} = LL^T$$

where $L \in R^{n \times n}$ is a lower triangular matrix.

The input variables $\mathbf{X} = [X_1, \cdots, X_n]$ can then be rewritten as

$$(X_1, \cdots, X_n)^T = (\mu_1, \cdots, \mu_n)^T + L \cdot (Z_1, \cdots, Z_n)^T$$

where $Z_j \sim N(0,1),\ j = 1, \cdots, n$ are independent standard normal variables.

The samples of Z_j, $j=1,\cdots,n$ can be generated independently by the process discussed above. The samples of $\mathbf{X}$ can then be calculated based on the samples of Z_j, $j=1,\cdots,n$.

When $\mathbf{X}=[X_1,\cdots,X_n]$ follows a general joint distribution other than joint normal, in principle the samples can be generated based on conditional distributions. It is however difficult to implement in practice for high dimensional input parameters. The joint normal transform method, called NORTA(NORmal To Anything), is a more realistic alternative when the uncertainty in $\mathbf{X}$ is modelled by the marginal distributions and the rank correlations [Kurowicka and Cooke 2006].

Suppose that the marginal distribution of X_i is F_i, $i=1,\cdots,n$, and the dependence among the X_i's is modelled by the rank correlation matrix $\mathbf{\Sigma}_X$. We need to generate the samples of the X_i's that satisfy both the marginals F_i and the $\mathbf{\Sigma}_X$. We start with defining a set of standard normal random variables $\mathbf{Y}=[Y_1,\cdots,Y_n]$, i.e. $Y_i \sim N(0,1)$, $i=1,\cdots,n$. Suppose the correlation matrix of $\mathbf{Y}$ is $\mathbf{\Sigma}_Y$. Consequently we have defined $\mathbf{Y} \sim N(\mathbf{0},\mathbf{\Sigma}_Y)$. Then the samples of $\mathbf{X}$ can be generated by:

- Generate the k^{th} sample of $\mathbf{Y} \sim N(\mathbf{0},\mathbf{\Sigma}_Y)$, denoted as $y_{k,1},\cdots,y_{k,n}$;
- Calculate the sample of X_i by $x_{k,i}=F_i^{-1}(\Phi(y_{k,i}))$, $i=1,\cdots,n$

The function $\Phi(\ )$ stands for the cumulative probability function of the standard normal distribution. The samples of $\mathbf{X}$ generated by this procedure will have the preset marginal distributions. To show this, note that each Y_i has a standard normal distribution, so that $\Phi(Y_i)$ is uniformly distributed on (0, 1). Therefore $X_i=F_i^{-1}(\Phi(Y_i))$ has the preset marginal distribution. The preset rank correlation matrix $\mathbf{\Sigma}_X$ can be obtained by setting the correlation matrix $\mathbf{\Sigma}_Y$ accordingly. However, there is no general closed-form expression that gives $\mathbf{\Sigma}_Y$ in terms of $\mathbf{\Sigma}_X$. Indeed, determining the right $\mathbf{\Sigma}_Y$ is perhaps the most difficult step in implementing

the NORTA method. More properties of the NORTA methods are investigated in [Ghosh and Henderson 2002; Ghosh and Henderson 2003]

Random sampling is easy to implement and provides unbiased estimates of the means, variance and the distribution of the model output. It suffers however large variance. In practice we need to generate a large number of samples to reduce the variance [Ghosh and Henderson 2003; Kurowicka and Cooke 2006]. It is therefore preferred when sufficient samples are possible and the computation of the model is not costly. As a complement, Latin hyper-cube sampling is developed for the situation when a large number of samples are impossible for computationally costly models.

2.6.3.3 Latin Hyper-cube Sampling

Latin hyper-cube sampling (LHS) is designed to ensure the full coverage of the input space [McKay and Beckman 1979] and therefore reduce the variability associated with the random sampling. Suppose that X has the CDF $F_X(x)$. To get n_s samples of X, the support of X is partitioned into n_s contiguous intervals χ_i, $i=1,\cdots,n_S$ of equal probability. Within χ_i, a sample x_i is generated by the random sampling method following the conditional distribution $F_{X|X\in\chi_i}(x)$. Following this process for χ_i, $i=1,\cdots,n_S$, we have n_S samples X.

For independent multiple input parameters $\mathbf{X}=[X_1,\cdots,X_n]$, the LHS samples can be generated for X_j, $j=1,\cdots,n$ separately at the first step. Denote the samples for X_j as $x_{i,j}$, $i=1,\cdots,n_S$. The LHS samples of $\mathbf{X}$ can then be generated by combining $x_{i_1,j}$ and $x_{i_2,k}$, $i_1,i_2=1,\cdots,n_S$, $j,k=1,\cdots,n$ randomly and without replacement [Helton and Davis 2000].

The LHS samples for dependent $\mathbf{X}=[X_1,\cdots,X_n]$ can be generated by the restricted pairing technique proposed in [Iman and Conover 1982]. The LHS samples are

generated for the input parameters separately at the first stage as for the independent input parameters. At the second step, the samples are combined together with specified selection procedure to obtain the preset rank correlations. A drawback to restricted pairing technique is that adding new samples to the existing samples will destroy the rank correlations in the samples. It means that the sample size cannot be increased by simply adding more samples as for the random sampling method. To solve this problem, an extension of the restricted pairing technique has been developed in [Sallaberry, Helton et al. 2006].

The LHS method can reduce the variance of the uncertainty analysis and so requires much less samples than the random sampling method [Stein 1987]. It however has poor coverage in the both ends of the input parameters. It is therefore preferable when a large number of samples are infeasible for computationally expensive models. Detailed comparisons between random sampling and LHS can be found in [McKay and Beckman 1979; Helton and Davis 2002; Helton and Davis 2003; Helton, Davis et al. 2005; Helton, Johnson et al. 2006].

2.6.4 Uncertainty Analysis of Fault-tree and Event-tree Models

For a Fault-tree model of independent basic events, the analytical solution of the variance of the top event is developed in [Rushdi 1985]. The analytical solution of the variance of the top event is also derived when the basic events follow a joint normal distribution [Der Kiureghian 1987].

For general applications, however, the uncertainty analysis with Fault-tree and Event-tree models is mainly conducted through simulations. Many examples of this type of uncertainty analysis can be found in the reactor safety field [Helton, Johnson et al. 1995; Helton, Johnson et al. 1995; Helton, Johnson et al. 1995; Helton, Anderson et al. 1996; Helton, Bean et al. 1996; Helton, Bean et al. 1997; Helton 1999; Helton, Anderson et al. 2000; Helton, Martell et al. 2000; Kraan and Cooke 2000]. More examples can also be found in chemical industry [Lauridsen, Kozine et al. 2002], offshore transport [Nilsen, Gudmestad et al. 1998], food safety [Frey and

Patil 2001], and natural disaster analysis [Iman, Johnson et al. 2002; Bazzurro and Luco 2005; Li and Ellingwood 2006] etc.

2.7 Conclusion

A good elicitation method should have strong foundations that are defensible in terms of probability and statistics theory. It should be as general as possible to apply in a wide variety of situations. It should also be able to be linked directly to the modelling procedure. The elicitation quantity should have a clear interpretation so that the assessors can view it as easy and credible [Clemen, Fischer et al. 2000]. Furthermore, for high dimensional input parameters, the method should require reasonable elicitation time from the experts. By these criteria, methods for eliciting the marginal probabilities are applicable in practice. The methods for eliciting the correlations between two random variables can also satisfy most of the applications. There is still a gap however in building high dimensional dependence by elicitation.

It has been agreed that uncertainty for high dimensional input parameters should be modelled on parametric marginal distributions plus a framework for modelling the correlations [Kurowicka and Cooke 2006; O'Hagan, Buck et al. 2006]. The correlation matrix or the rank correlation matrix offers the straightforward framework that is easy to understand and can be dealt with by the well developed algorithms for matrices. It however requires assessing $n(n-1)/2$ pairwise correlations for the n input parameters, which means too much elicitation workload when n is large. Furthermore, by eliciting pairwise correlations it is very difficult to keep the outcome matrix positive definite or positive semi-definite as required for a correlation matrix. Copula trees offer an efficient and consistent framework when the input parameters hold suitable structure patterns [Bedford and Cooke 2001; Bedford and Cooke 2002]. Copula vines can be used for all kinds of input parameters. The elicitation for populating the model is quite difficult to implement in practice. The method for building the dependence on the factors is an efficient and flexible way that also offers a structured exploration of the underlying dependence causes [Van Dorp 2005]. No correlation or covariance matrix however has been developed with

the method. Therefore it cannot be applied when we do need a covariance matrix for modelling the uncertainty in the input parameters.

The uncertainty analysis of Fault-tree and Event tree models can be efficiently conducted by simulations-based methods in conjunction with analytical analysis of the moments. In the literature, however, there are still no methods found for calculating the mean and the variance when both the basic events of the Fault-tree models and the consequences of the Event tree models are treated as correlated random variables. As we have discussed in Chapter 1, fault trees and event trees are usually built with commercial software packages such as *Isograph FT+* and stored as computer codes in the internal database. For uncertainty analysis, we need to access these computer models. In *Isograph FT+*, an interface is developed for users to set the input parameters, to run the fault trees and event trees, and to obtain the results. By this interface, we can conduct simulations of the fault trees and event trees for uncertainty analysis. However, the interface is developed as dynamic link libraries (DLLs). It requires high level of programming skills to utilize this interface. Furthermore, we cannot do any analytical analysis of the uncertainty of fault trees and event trees though that interface. There is still no discussion on how to implement the analysis on a large model built with a commercial software package.

In this research, we aim at first to develop a method for building the correlation matrix through elicitation on uncertainty factors. The method is expected to have two features: (1) significant reduction of the elicitation workload compared with direct elicitation of the correlation matrix; (2) guaranteed positive definite or positive semi-definite correlation matrix. Secondly, we aim to develop a method to reduce the elicitation for deriving the variance of the input parameters. Thirdly we aim to develop a method to model the dependence among the rule sets associated with the consequences of the Event-tree models. Fourthly, we aim to select or develop a set of methods for conducting the uncertainty analysis of the Fault-tree and Event-tree models built on the software package *Isograph FT+*.

Chapter 3

A New Procedure for Building the Covariance Matrix of Input Parameters through Elicitation

3.1 Overview

For a large group of input parameters, it is difficult to build the uncertainty model through elicitation of expert judgment. In practice, it is split into two tasks: (1) building the marginal distributions for all the individual parameters; and (2) building the dependence among all the parameters [Kurowicka and Cooke 2006]. When the normal distribution is assumed as the marginal for each single parameter, the uncertainty in a group of input parameters can be completely defined by the means and the covariance matrix. In this chapter, a procedure is developed for building the covariance matrix through the elicitation of expert judgement.

Uncertainty factors have been employed as an efficient way in modelling the uncertainty in the input parameters. The elicitation workload can be reduced significantly by modelling the uncertainty factors [Van Dorp 2005]. In this chapter, the uncertainty factors are used to structure the input parameters' uncertainty through a linear model [Cheng, Bedford et al. 2007]. The correlation matrix of the input parameters is then developed based on the elicitation of the correlations regarding the uncertainty factors. The outcome correlation matrix can be guaranteed to be positive semi-definite as required.

We continue then to develop the method for deriving the variance of the input parameters through expert judgement elicitation. The input parameters that are

defined physically in a similar way are defined to be a family. When input parameters are occurrence rates of a Poisson process, the sum of the family members is the occurrence rate of the family. The variance of the family occurrence rate, called the variance of the family, is then meaningful and can be derived by eliciting the percentile of the waiting time [Garthwaite, Kadane et al. 2005]. A method is then developed to derive the family covariance matrix based on the family variance and the family correlation matrix. This method at first features a significant reduction of the elicitation workload. Furthermore some experts may be more confident in giving their assessment on higher level events, as was the case with the experts in RSSB in the case studies for this research.

To implement the above method for building the correlation matrix, the correlation between two random variables must be elicited. The qualitative assessment plus benchmark method is adopted because it is easy for experts to make the judgement [Clemen, Fischer et al. 2000] and has been widely used in the Excel Add-in *Crystal Ball* [ORACLE 2008]. The method is presented and discussed later in this chapter.

A method is then developed for deriving the family variance based on the elicitation of the percentile of the waiting time for the next occurrence.

At the end of this chapter, we include an example for demonstrating the procedure for building the covariance matrix of an input parameter family.

3.2 Research philosophy and Methodology

3.2.1 Introduction

This research focuses on uncertainty analysis of a risk assessment model. The uncertainty can arise from two perspectives. First, a model is always an approximate representation of reality [Pidd 2003]. Consequently when assessment is made based on a model, uncertainty is caused by the unknown approximation. This uncertainty is called model uncertainty. Second, with a given model, we usually cannot know for sure the values of the input parameters. The uncertainty in the input parameters

propagates through the model and causes the uncertainty in the output. This is called parameter uncertainty.

This research focuses on the parameter uncertainty analysis. We need to model the uncertainty in the parameters at first and then assess the resulting in uncertainty in the output. To quantitatively assess the uncertainty, we must define a mathematical representation of it. Probability theory offers a rigorous framework for this and has been well accepted in the literature [Lindley 2000]. Kolmogorov's axioms states that probability is a positive normalized measure over a field of "possible worlds" or "possible states of nature". This axioms has been widely accepted as the most appropriate framework for uncertainty assessment. Within the field of uncertainty of PRA, however, strong debate exists over the actual meaning of probability [Apolostolakis 1988; Winkler 1996; Nilsen and Aven 2003]. There are two main interpretations of probability: the frequentist and subjectivist views. The choice of the interpretation depends on the way the researcher observes the world, and acts within it. Consequently it defines the methodology: i.e. the ways for date collection and the data analysis. We are going to discuss the philosophy standpoint and justify the adopted methodology regarding this research. Specifically we will discuss three questions: the ontology, the epistemology and the methodology.

3.2.2 Ontology

Ontology concerns the form and nature of reality. Positivists believe that there is only one true reality; interpretivists accept that the reality depends on one's experience and perceptions and therefore is subjective [Ponterotto 2005]. Frequentists and subjectivests of probability hold completely different ontological positions. On one hand, frequentists believe that a true reality exists independent to observers. To a frequentist, probability is a natural characteristic of the world. For a specific event, there is a true probability. On the other hand, subjectivists believe that probability is a degree of belief in the occurrence of an event that is decided by the observer's knowledge [De Finetti 1974]. Correspondingly, subjective probability is a mode of judgment made by an observer based on his/her knowledge and background information. As a result, uncertainty is not an objective characteristic of an event

existing in the real world independently of the analyst as frequentists believe [Parry 1996].

As stated before, this research aims to assess the uncertainty in the output driven by the uncertainty in the input parameters. We study the epistemic uncertainty that represents the lack of the experts' knowledge. Different experts may have different knowledge about the input parameters that result in different definitions of the probabilities. Therefore we adopt the subjectivist view of probability; correspondingly we take the interpretivists' view of ontology.

3.2.3 Epistemology

Epistemology concerns the process of knowledge acquisition and defines the relationship between the knower and what can be known.

For a frequentist, probability learning is integrally related to the concept of relative frequencies. Because probability is believed as a characteristic of the objective world, in order to assess this probability, one needs to perform a series of repeatable experiments in which this event occurs. The probability of the event is then set as the relative frequency of the outcomes of this event. When the number of the experiments is large, the relative frequency converges to the true probability of the event. Mathematically, we can define a sequence of independent random variables $\{X_1, \cdots, X_n, \cdots\}$ and X_i is the outcome of the *i*th experiment defined as

$$X_i = \begin{cases} 1, & \text{the event happens} \\ 0, & \text{otherwise} \end{cases}$$

The probability of the event can then be inferred by

$$P = \lim_{n \to +\infty} \frac{\sum_{i=1}^{n} X_i}{n}$$

Associated with point estimation of the probability, a confidence interval can also be estimated representing a level of acceptability for the estimation. As a summary frequentists' epistemology is to use the historical data to estimate probability that is believed to be an objective true value.

For an interpretivist, the probability represents one's degree of belief of an event of interest. In order to assess the probability we need to explore one's knowledge and beliefs and formulate them into probability. This process is commonly called expert judgement elicitation [Garthwaite, Kadane et al. 2005; O'Hagan, Buck et al. 2006]. Expert judgement elicitation has been used in many fields including risk and uncertainty assessment for decades [Cooke 1991]. With continuous effort in improving the elicitation, it is increasingly employed for uncertainty assessment [Kurowicka and Cooke 2006; O'Hagan, Buck et al. 2006].

A sound process is essential to implement a good elicitation. Broadly an elicitation process covers four basic stages: Set up, Elicit, Fit, and Evaluate [Garthwaite, Kadane et al. 2005]. In the set up stage, we select the experts, training the experts, identifying what aspects of the problem to elicit. The Elicit stage is the implementation of elicitation. In the Fit stage the experts' answers are transferred into the mathematical models to derive probabilities. In the Evaluate stage, we evaluate whether the elicitation outcome is adequate or not. We need to emphasize that an elicitation is well done if the elicited information accurately represents the expert's belief and knowledge. It is not related to how good that knowledge is [Cooke 1991; Garthwaite, Kadane et al. 2005].

3.2.4 Methodology

Methodology concerns how the data are obtained in a research. It depends on the ontology and epistemology adopted for the research. From the frequentist's view of probability, we need occurrence data of an event to estimate its probability as discussed above. To obtain data, one way is to observe the event and record the historical occurrences. Another way is to conduct controlled experiments in a sufficiently identical way. It is sometimes a faster and cheaper way to obtain data.

From the subjectivist's view of probability, as stated before, we elicit information from experts and then formulate the information into probabilities. Therefore elicitation is the main way for subjectivist's to obtain data. It can be implemented through questionnaires, interviews, workshop etc [Creswell 2009]. One important task for conducting elicitation is to design a sound mathematical framework for formulating the elicited results. It is intrinsically related to the questions that we used to elicit experts' knowledge and beliefs [Cooke 1991; Bedford and Cooke 2001; O'Hagan, Buck et al. 2006].

As stated before, in this research we take the subjectivist's interpretation of probability in line with the interpretivists' view of ontology. Correspondingly we adopt expert judgement elicitation to evaluate probabilities for uncertainty assessment. We first chose and develop proper mathematical frameworks for information fusion. Accordingly we design questions for experts to answer.

Experts are chosen from RSSB for three reasons. First, it is the RSSB that builds the safety risk model. The RSSB now is concerned with their uncertainty in the model. The experts are likely to have a positive attitude towards this exercise. Second, the experts know how the RSSB-SRM is built and how it is used in practice. As such, they are knowledgeable about the uncertainty sources. Third, the experts have at least basic understanding of statistics. Therefore they can understand well elicitation questions. Before starting elicitation, training was given for experts to understand the mathematical framework. Documents regarding the mathematical framework were first offered to experts. After experts finished reading, a two-day's workshop was organized and a mock elicitation was conducted to make sure that experts gained understanding of the relevant issues.

In this research, we conducted two case studies. For Case I, elicitation was implemented through questionnaires. Four forms were designed for benchmark, family identification, family definitions respectively. Eight families were identified by experts for Case I. For each family, three forms were designed to elicit

information regarding the uncertainty factors. The whole elicitation lasted for 2 months as experts worked part-timely for the elicitation. During the elicitation process, the researcher and the experts communicated through both email and phone talk. Case study II was conducted through a five-day's workshop. The researcher and the experts worked together to elicit the information. It turns out that face-to-face workshop is a more productive way to eliciting.

Following the subjectivist's view of probability, there is not a real probability that can be used to validate the elicited information. A good elicitation is done when it catches properly the knowledge and beliefs of experts [Cooke 1991; Garthwaite, Kadane et al. 2005]. Accordingly, the validation was conducted at two stages in this research. First, we derive an indicator to monitor the elicited information during the elicitation process. The information given by experts can be guaranteed to be consistent once the indicator is satisfied. The case studies demonstrate that the indicator works very efficiently. Second, we need to make sure that the elicited information represents the knowledge of all experts in the institute, rather than the individual experts participating in the elicitation. For this end, the experts participating in the elicitation were asked to communicate with the experts in the institute during the elicitation process. At the last step of the elicitation, a review meeting was arranged for the relevant experts to check the elicited information and the derived results. The case studies demonstrate that the review meeting is an efficient way.

3.3 Linear Model on Uncertainty Factors

Assume that the uncertainty in a parameter is affected by a set of independent factors. Assume further that the uncertainty is defined with a linear model of the factors as

$$p_i - \mu_i = \lambda_{i,1} X_{i,1} + \lambda_{i,2} X_{i,2} + \cdots + \lambda_{i,K_i} X_{i,K_i} + w_i \varepsilon_i \tag{3.1}$$

where p_i stands for an input parameter; μ_i is the mean of p_i; $X_{i,k}$ is the k^{th} factor that is normalized such as $\mathrm{var}(X_{i,k}) = 1$ and $E(X_{i,k}) = 0$; $\lambda_{i,k}$ is the loading of $X_{i,k}$;

K_i is the number of the factors of p_i; ε_i stand for the residual uncertainty of p_i, which is normalized such as $\mathrm{var}(\varepsilon_i)=1$, $E(\varepsilon_i)=0$; ε_i is assumed to be independent to all the factors, i.e. $\mathrm{cov}(X_{i,k},\varepsilon_i)=0$, $k=1,\cdots,K_i$; w_i is the loading of ε_i.

We assume that p_i and $X_{i,k}$ in Equation 3.1 follow a normal distribution. Rather than for prediction of p_i, the linear model of Equation 3.1 aims to structure the uncertainty of the input parameter into the factors. Correlations between two parameters can then be derived through the shared uncertainty factors.

Based on Equation 3.1, the variance of p_i can be derived as

$$\mathrm{var}(p_i)=\sum_k \lambda_{i,k}^2 + w_i^2 \tag{3.2}$$

Similarly if we can define another parameter p_j in the same form as in Equation 3.1, then the covariance between p_i and p_j is

$$\mathrm{cov}(p_i,p_j)=\sum_{\ell=1}^{K_i}\sum_{k=1}^{K_j}\lambda_{i,\ell}\lambda_{j,k}\cdot r(X_{i,\ell},X_{j,k})+w_i w_j\,\mathrm{cov}(\varepsilon_i,\varepsilon_j) \tag{3.3}$$

where $r(,)$ stands for the correlation between the two variables.

The correlation $r(X_{i,\ell},X_{j,k})$ is included in Equations 3.3 because the factors from two different parameters could be identical or correlated while the factors associated with one parameter are assumed to be independent. By rearranging the sequence of the factors and adding zero coefficients $\lambda_{i,k}$ when needed, the relations among the factors are demonstrated in Fig. 3.1. In a horizontal oval, the factors $X_{i,k}$, $k=1,\cdots,K$, $K=\max_i\{K_i\}$, are associated with p_i and are assumed to be independent. In a vertical oval, the factors $X_{i,k}$ and $X_{j,k}$ can be correlated or even

identical. Consequently, two factor $X_{i,k(1)}$ and $X_{j,k(2)}$, where $k(1),k(2)=1,\cdots,K$, $k(1)\neq k(2)$, are independent. We will discuss the classifications of the factors so that suitable definitions can be made accordingly.

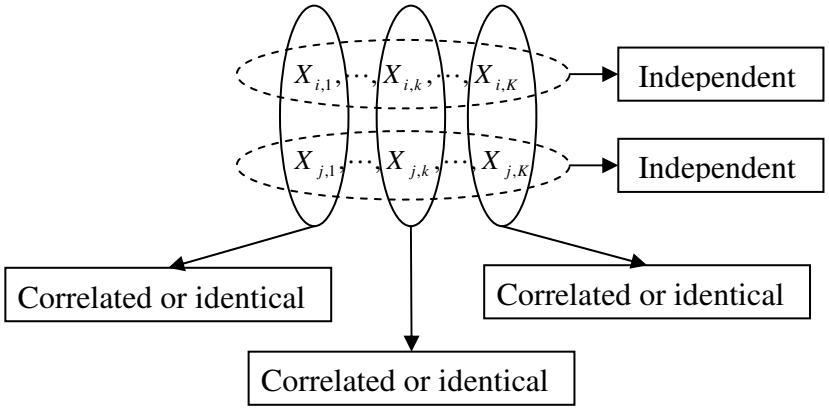


Fig. 3.1 Relations among the factors associated with multiple parameters

3.4 Factor Categories

For deriving the correlation between the input parameters, the factors associated with p_i and p_j are partitioned into three categories: common factors, correlated factors and exclusive factors, as demonstrated in Fig. 3.2.

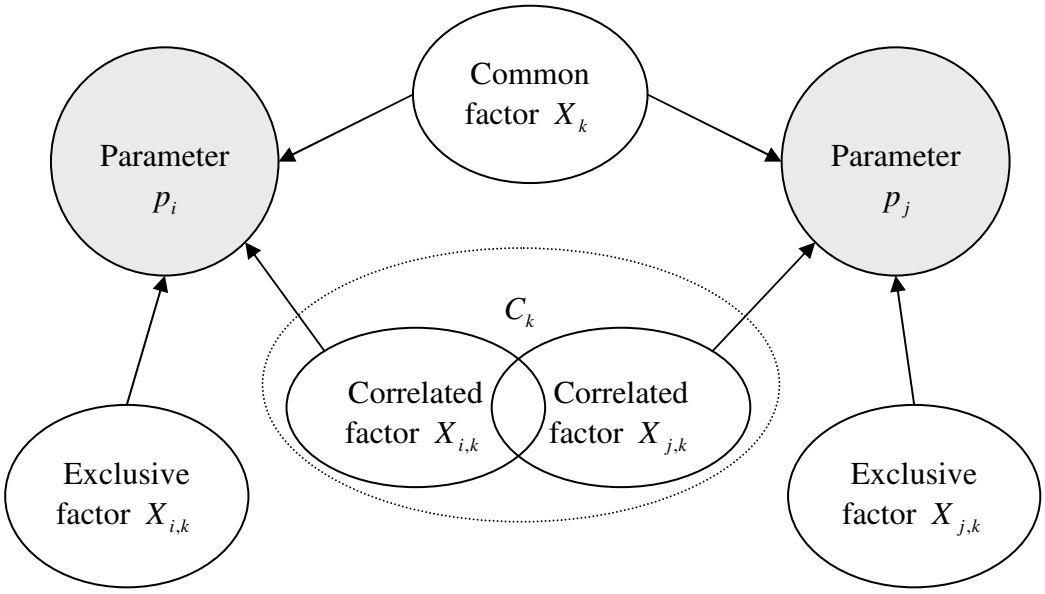


Fig. 3.2 Classifications of the uncertainty factors

As shown in Fig. 3.1, suppose that $X_{i,k}$ and $X_{j,k}$ are two factors from p_i and p_j respectively. When $\lambda_{i,k} \neq 0$, $\lambda_{j,k} \neq 0$, $X_{i,k}$ and $X_{j,k}$ define a common factor when they are identical. It is denoted by $X_k = X_{i,k} = X_{j,k}$.

When $\lambda_{i,k} \neq 0$, $\lambda_{j,k} \neq 0$, $X_{i,k}$ and $X_{j,k}$ define a factor class when they are correlated. It is denoted by $C_k = \langle X_{i,k}, X_{j,k} \rangle$.

When $\lambda_{j,k} = 0$, $X_{i,k}$ is associated with p_i and has nothing to do with p_j. Consequently $X_{i,k}$ is called exclusive factor of p_i. Accordingly, when $\lambda_{i,k} = 0$, $X_{j,k}$ is an exclusive factor of p_j.

To exemplify the above classifications, we take a group of RSSB-SRM precursors that are defined as *road vehicle (RV) driver error causing RV struck by train on level crossing L,* where *L* denotes one of the eight types of level crossings (LCs) [RSSB 2004]. These precursors are defined as a collective event for all the level crossings of the same type. The eight types of LCs are homogenously deployed across the Britain railway mainline [Dennis 2006]. The eight precursors are then exposed homogenously to the weather over Britain. We think bad weather condition can cause more road vehicle drivers' errors causing RV struck by train on level crossings. When the weather over Britain deviates consistently from the historical average, the eight precursors are expected to have a deviation of the occurrence rates consistently. Because weather affects the eight precursors in the same way, it is a common factor for all the eight precursors.

One of the eight level crossings is the User-Worked Crossing with Telephone (UWC-T). The user is required to use the phone to call the signaller in order to obtain permission to use the crossing. When the crossing is clear for railway use, the user is required to use the phone to inform the signaller. If the reliability of the phone-

communication system is considered as a factor, it is exclusive to UWC-T because no other level crossings are equipped with it.

The eight types of LCs can be put into two groups defined as: User Worked Crossings (UWCs) and non-UWCs. The level crossings in each group have the same user populations and the two populations are different to some extent [RSSB 2004]. Therefore, two factors can be defined according to the two groups of usage population respectively. The two factors are correlated because there is still an overlap of the two user populations. Based on the above definition, these two factors form a factor class. As we can see, the factor class actually composes of the two correlated factors, each of which is a common factor to a subset of the eight precursors. To populate the model, we need to assess only one correlation for this correlated factor class. Such a class is called thin class. A thin class makes it feasible to model the correlated factors by elicitation. In practice, we can always keep a factor class thin by agglomerating some correlated factors in the class.

According to the above classifications, the input parameter's linear model defined in Equation 3.1 can be rewritten as

$$\begin{aligned} p_i - \mu_i &= \sum_{k \in L_X} \lambda_{i,k} X_k + \sum_{k \in L_C} \lambda_{i,k} X_{i,k} + \sum_{k \in L_{E,i}} \lambda_{i,k} X_{i,k} + w_i \varepsilon_i \\ p_j - \mu_j &= \sum_{k \in L_X} \lambda_{j,k} X_k + \sum_{k \in L_C} \lambda_{j,k} X_{j,k} + \sum_{k \in L_{E,j}} \lambda_{j,k} X_{j,k} + w_j \varepsilon_j \end{aligned} \tag{3.4}$$

where L_X , L_C stand for the sets of indices for the common factors and the correlated factors respectively; $L_{E,i}$ and $L_{E,j}$ stand for the sets of indices for the exclusive factors associated with p_i and p_j respectively.

Any two input parameters are connected through their common factors and the factor classes. In the next section, the correlation between two parameters is to be derived based on the model defined in Equation 3.4.

3.5 Correlation between Two Input Parameters

We assume that the residual uncertainty of different parameters are independent i.e., $\mathrm{cov}(\varepsilon_i, \varepsilon_j) = 0$. According to the classifications of the factors, and recalling that all the factors associated with one input parameter are independent, the covariance between p_i and p_j can be derived from Equations 3.3 and 3.4 as

$$\mathrm{cov}(p_i, p_j) = \sum_{k \in L_X} \lambda_{i,k} \lambda_{j,k} + \sum_{k \in L_C} \lambda_{i,k} \lambda_{j,k} \cdot r(X_{i,k}, X_{j,k}) \tag{3.5}$$

For subjective uncertainty analysis, we rely on the elicitation of the expert judgement to assess the loadings $\lambda_{i,k}$. It is however difficult for experts to assess $\lambda_{i,k}$ because it involves both the correlation and the variance [Garthwaite, Kadane et al. 2005; O'Hagan, Buck et al. 2006]. Consequently the covariance in Equation 3.5 cannot be derived directly from the loadings. As an alternative, the correlation between p_i and p_j is to be derived based on elicitation of the correlations between the factors. The variance of the parameter is to be derived through elicitation separately.

For a factor $X_{i,k}$ of p_i, we have

$$\lambda_{i,k} = \frac{\mathrm{cov}(p_i, X_{i,k})}{\mathrm{var}(X_{i,k})} = \frac{\mathrm{cov}(p_i, X_{i,k})}{\sqrt{\mathrm{var}(X_{i,k})}\sqrt{\mathrm{var}(p_i)}} \frac{\sqrt{\mathrm{var}(p_i)}}{\sqrt{\mathrm{var}(X_{i,k})}} = r(p_i, X_{i,k}) \frac{\sqrt{\mathrm{var}(p_i)}}{\sqrt{\mathrm{var}(X_{i,k})}}$$

Recalling that $\mathrm{var}(X_{i,k}) = 1$, from the above formula we have

$$\lambda_{i,k} = r(p_i, X_{i,k}) \sqrt{\mathrm{var}(p_i)} \tag{3.6}$$

Applying Equation 3.6 with Equation 3.5 we obtain

$$\operatorname{cov}(p_i, p_j) = \left(\sum_{k \in L_X} h_{i,j}^k + \sum_{k \in L_C} o_{i,j}^k \right) \cdot \sqrt{\operatorname{var}(p_i)\operatorname{var}(p_j)} \tag{3.7}$$

where

$$h_{i,j}^k = r(p_i, X_k) \cdot r(p_j, X_k) \tag{3.8}$$

$$o_{i,j}^k = r(p_i, X_{i,k}) \cdot r(p_j, X_{j,k}) \cdot r(X_{i,k}, X_{j,k}) \tag{3.9}$$

By the definition of correlation, we have

$$r(p_i, p_j) = \frac{\operatorname{cov}(p_i, p_j)}{\sqrt{\operatorname{var}(p_i)\operatorname{var}(p_j)}} = \sum_{k \in L_X} h_{i,j}^k + \sum_{k \in L_C} o_{i,j}^k \tag{3.10}$$

In Equation 3.10, the term $h_{i,j}^k$ and $o_{i,j}^k$ are the contributions from the common factors and factor classes respectively. The calculation of $h_{i,j}^k$ and $o_{i,j}^k$ is based on the correlations regarding the factors, which can be reasonably elicited as well discussed in the literature [Clemen, Fischer et al. 2000; Garthwaite, Kadane et al. 2005; O'Hagan, Buck et al. 2006]. The correlation coefficients that need to elicit are summarized as:

- $r(p_i, X_k)$: the correlation between a common factor and each input parameter
- $r(p_i, X_{i,k})$: the correlation between a classed factor and its associated input parameter
- $r(X_{i,k}, X_{j,k})$: the correlation between two factors of the same class

We have assumed that all the factors associated with the same input parameter are independent, therefore are uncorrelated; correspondingly no correlations among them need to elicit. Based on the assumption, two classed factors $X_{i,k}$ and $X_{j,k}$ are

uncorrelated with all the other factors $X_{n,m}$, $m \neq k$. It does not however mean necessarily that $X_{i,k}$ and $X_{j,k}$ are uncorrelated. Their correlation needs to be elicited.

Through the above method, a positive semi-definite correlation matrix can be guaranteed for multiple parameters as will be shown in the next section.

3.6 Correlation Matrix for Multiple Parameters

3.6.1 Overview

Suppose that there are *N* parameters arranged in a vertical vector as defined as

$$\mathbf{p} = [p_1, \cdots, p_n, \cdots, p_N]^T \tag{3.11}$$

The correlation matrix of **p** is denoted by

$$r(\mathbf{p}) = \begin{bmatrix} 1 & \cdots & r(p_1, p_n) & \cdots & r(p_1, p_N) \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ r(p_n, p_1) & \cdots & 1 & \cdots & r(p_n, p_N) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r(p_N, p_1) & \cdots & r(p_N, p_n) & \cdots & 1 \end{bmatrix} \tag{3.12}$$

The matrix $r(\mathbf{p})$ can be built with each cell $r(p_i, p_j)$ calculated separately from Equation 3.10. We are however going to define $r(\mathbf{p})$ in a compact form as the sum of a series of matrices associated with the common factors, correlated factors, the exclusive factors and the residual terms. We will prove that those summand matrices can be kept positive semi-definite and as a result $r(\mathbf{p})$ is positive semi-definite. For this end, related definitions and theories on positive definite or positive semi-definite matrices are included in Appendix C.

3.6.2 Decomposition of the Correlation Matrix of Multiple Input Parameters

For developing the correlation matrix of $\mathbf{p}$ as denoted in Equation 3.11, we are going to redefine the linear uncertainty model in Equation 3.4 in a compact format with matrices. For this purpose, we need to define the column vectors of the factors associated with the common factors, the factor classes, and the exclusive factors as defined before. We denote a column vector of the factors as $\mathbf{X}_k$, $\mathbf{X}_k \in \mathbb{R}^{N\times 1}$, where N is the number of the input parameters included in $\mathbf{p}$. Associated with each $\mathbf{X}_k$, we need to define the matrices of the loadings, which is denoted by $\boldsymbol{\lambda}_k$, $\boldsymbol{\lambda}_k \in \mathbb{R}^{N\times N}$.

Associated with a common factor, X_k, $k \in L_X$, we define the column vector of factors as

$$\mathbf{X}_k = \left(X_k, \cdots, X_k\right)^T \tag{3.13}$$

Accordingly we define the matrix of loadings as

$$\boldsymbol{\lambda}_k = \begin{bmatrix} \lambda_{1,k} & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & \lambda_{i,k} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & \lambda_{N,k} \end{bmatrix} \tag{3.14}$$

Therefore $\boldsymbol{\lambda}_k$ is a diagonal matrix with all the off-diagonal cells filled with zeros.

Associated with a factor class C_k, $k \in L_C$, we define the column vector of factors as

$$\mathbf{X}_k = \left(X_{1,k}, \cdots, X_{i,k}, \cdots, X_{N,k}\right)^T \tag{3.15}$$

Accordingly we define the matrix of loadings as

$$\boldsymbol{\lambda}_k = \begin{bmatrix} \lambda_{1,k} & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \cdots & \vdots \\ 0 & \cdots & \lambda_{i,k} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & \lambda_{N,k} \end{bmatrix} \tag{3.16}$$

Hence $\boldsymbol{\lambda}_k$ is a diagonal matrix with all the off-diagonal cells filled with zeros.

Associated with an exclusive factor $X_{i,k}$, $k \in L_{E,i}$, we define the column vector of factors as

$$\mathbf{X}_k = \left(X_{i,k}, \cdots, X_{i,k}\right)^T \tag{3.17}$$

Accordingly we define the matrix of loadings as

$$\boldsymbol{\lambda}_k = \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \cdots & \vdots \\ 0 & \cdots & \lambda_{i,k} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} \tag{3.18}$$

The matrix $\boldsymbol{\lambda}_k$ in Equation 3.18 contains only one non-zero value on the i^{th} diagonal cell.

We define the column vector associated with the residual ε_i, $i = 1, \cdots, N$, as

$$\boldsymbol{\varepsilon}_i = \left(\varepsilon_i, \cdots, \varepsilon_i\right)^T,\ \boldsymbol{\varepsilon}_i \in \mathbb{R}^{N \times 1} \tag{3.19}$$

Accordingly we define the matrix of loadings associated with $\boldsymbol{\varepsilon}_i$ as

$$\mathbf{w}_i = \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \cdots & \vdots \\ 0 & \cdots & w_i & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{bmatrix},\ \mathbf{w}_i \in \mathbb{R}^{N\times 1} \tag{3.20}$$

The matrix $\mathbf{w}_i$ contains only one non-zero value on the i^{th} diagonal cell.

Based on the definitions in Equations 3.13-3.20 and the linear model in Equation 3.4, we can define the linear model of the *N* input parameters in terms of the matrices as

$$\mathbf{p}-\mathbf{u} = \sum_{k\in L_X} \boldsymbol{\lambda}_k \mathbf{X}_k + \sum_{k\in L_C} \boldsymbol{\lambda}_k \mathbf{X}_k + \sum_{i=1}^{N} \sum_{k\in L_{E,i}} \boldsymbol{\lambda}_k \mathbf{X}_k + \sum_{i=1}^{N} \mathbf{w}_i \boldsymbol{\varepsilon}_i \tag{3.21}$$

Recall that we assume that the factors, except of the factors of the same class, and the residuals are mutually independent. As given in [Van Kampen 1992], we have

$$\begin{aligned} \operatorname{cov}(\mathbf{p},\mathbf{p}) = & \sum_{k\in L_X} \operatorname{cov}(\boldsymbol{\lambda}_k \mathbf{X}_k, \boldsymbol{\lambda}_k \mathbf{X}_k) + \sum_{k\in L_C} \operatorname{cov}(\boldsymbol{\lambda}_k \mathbf{X}_k, \boldsymbol{\lambda}_k \mathbf{X}_k) \\ & + \sum_{i=1}^{N} \sum_{k\in L_{E,i}} \operatorname{cov}(\boldsymbol{\lambda}_k \mathbf{X}_k, \boldsymbol{\lambda}_k \mathbf{X}_k) + \sum_{i=1}^{N} \operatorname{cov}(\mathbf{w}_i \boldsymbol{\varepsilon}_i, \mathbf{w}_i \boldsymbol{\varepsilon}_i) \end{aligned} \tag{3.22}$$

Suppose that $\mathbf{V} = (V_1, \cdots, V_N)^T$, $\mathbf{V} \in \mathbb{R}^{N\times 1}$, $\mathbf{U} \in \mathbb{R}^{M\times N}$, as given in [Dillon and Goldstein 1984] we have

$$\operatorname{cov}(\mathbf{UV}, \mathbf{UV}) = \mathbf{U} \operatorname{cov}(\mathbf{V}, \mathbf{V}) \mathbf{U}^T \tag{3.23}$$

Therefore from Equation 3.10 we have

$$\begin{aligned} \operatorname{cov}(\mathbf{p},\mathbf{p}) = & \sum_{k\in L_X} \boldsymbol{\lambda}_k \operatorname{cov}(\mathbf{X}_k, \mathbf{X}_k) \boldsymbol{\lambda}_k^T + \sum_{k\in L_C} \boldsymbol{\lambda}_k \operatorname{cov}(\mathbf{X}_k, \mathbf{X}_k) \boldsymbol{\lambda}_k^T \\ & + \sum_{i=1}^{N} \sum_{k\in L_{E,i}} \boldsymbol{\lambda}_k \operatorname{cov}(\mathbf{X}_k, \mathbf{X}_k) \boldsymbol{\lambda}_k^T + \sum_{i=1}^{N} \mathbf{w}_i \operatorname{cov}(\boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon}_i) \mathbf{w}_i^T \end{aligned} \tag{3.24}$$

We can derive the covariance matrix $\text{cov}(\mathbf{X}_k, \mathbf{X}_k)$ following the definitions of $\mathbf{X}_k$. When $\mathbf{X}_k$ is associated with a common factor, i.e. $k \in L_X$, from Equation 3.13 we have

$$\text{cov}(\mathbf{X}_k, \mathbf{X}_k) = \begin{bmatrix} \text{cov}(X_k, X_k) & \cdots & \text{cov}(X_k, X_k) & \cdots & \text{cov}(X_k, X_k) \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \text{cov}(X_k, X_k) & \cdots & \text{cov}(X_k, X_k) & \cdots & \text{cov}(X_k, X_k) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \text{cov}(X_k, X_k) & \cdots & \text{cov}(X_k, X_k) & \cdots & \text{cov}(X_k, X_k) \end{bmatrix}$$

Recalling $\text{cov}(X_k, X_k) = \text{var}(X_k) = 1$, we have

$$\text{cov}(\mathbf{X}_k, \mathbf{X}_k) = \begin{bmatrix} 1 & \cdots & 1 & \cdots & 1 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 1 & \cdots & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \cdots & 1 & \cdots & 1 \end{bmatrix} = [1]_{N \times N}, \; k \in L_X \tag{3.25}$$

where $[1]_{N \times N}$ is called ones-matrix of dimensions $N \times N$.

When $\mathbf{X}_k$ is associated with a factor class C_k, $k \in L_C$, from Equation 3.15, we have

$$\text{cov}(\mathbf{X}_k, \mathbf{X}_k) = \begin{bmatrix} \text{cov}(X_{1,k}, X_{1,k}) & \cdots & \text{cov}(X_{1,k}, X_{i,k}) & \cdots & \text{cov}(X_{1,k}, X_{j,k}) & \cdots & \text{cov}(X_{i,k}, X_{N,k}) \\ \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \vdots \\ \text{cov}(X_{i,k}, X_{1,k}) & \cdots & \text{cov}(X_{i,k}, X_{i,k}) & \cdots & \text{cov}(X_{i,k}, X_{j,k}) & \cdots & \text{cov}(X_{i,k}, X_{N,k}) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \text{cov}(X_{j,k}, X_{1,k}) & \cdots & \text{cov}(X_{j,k}, X_{i,k}) & \cdots & \text{cov}(X_{j,k}, X_{j,k}) & \cdots & \text{cov}(X_{j,k}, X_{N,k}) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \text{cov}(X_{N,k}, X_{1,k}) & \cdots & \text{cov}(X_{N,k}, X_{i,k}) & \cdots & \text{cov}(X_{N,k}, X_{j,k}) & \cdots & \text{cov}(X_{N,k}, X_{N,k}) \end{bmatrix}$$

(3.26)

We define

$$\sigma_{i,j}^2 = \mathrm{cov}\left(X_{i,k}, X_{j,k}\right) \tag{3.27}$$

Recalling $\mathrm{var}\left(X_{i,k}\right) = 1$, we have

$$\sigma_{i,j}^2 = r\left(X_{i,k}, X_{j,k}\right)\sqrt{\mathrm{var}\left(X_{i,k}\right)\cdot \mathrm{var}\left(X_{j,k}\right)} = r\left(X_{i,k}, X_{j,k}\right) \tag{3.28}$$

and

$$\sigma_{i,i}^2 = \mathrm{cov}\left(X_{i,k}, X_{i,k}\right) = 1 \tag{3.29}$$

Applying Equations 3.27 - 3.29 with Equation 3.26, we have

$$\mathrm{cov}\left(\mathbf{X}_k, \mathbf{X}_k\right) = \begin{bmatrix} 1 & \cdots & r\left(X_{1,k}, X_{i,k}\right) & \cdots & r\left(X_{1,k}, X_{j,k}\right) & \cdots & r\left(X_{1,k}, X_{N,k}\right) \\ \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \vdots \\ r\left(X_{i,k}, X_{1,k}\right) & \cdots & 1 & \cdots & r\left(X_{i,k}, X_{j,k}\right) & \cdots & r\left(X_{i,k}, X_{N,k}\right) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ r\left(X_{j,k}, X_{1,k}\right) & \cdots & r\left(X_{j,k}, X_{i,k}\right) & \cdots & 1 & \cdots & r\left(X_{j,k}, X_{N,k}\right) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ r\left(X_{N,k}, X_{1,k}\right) & \cdots & r\left(X_{N,k}, X_{i,k}\right) & \cdots & r\left(X_{N,k}, X_{j,k}\right) & \cdots & 1 \end{bmatrix}$$

$$\equiv r\left(C_k\right)$$

(3.30)

When $\mathbf{X}_k$ is associated with an exclusive factor, i.e. $k \in L_{E,i}$, from Equation 3.17 we have

$$\operatorname{cov}(\mathbf{X}_k,\mathbf{X}_k)=\begin{bmatrix}\operatorname{cov}(X_{i,k},X_{i,k}) & \cdots & \operatorname{cov}(X_{i,k},X_{i,k}) & \cdots & \operatorname{cov}(X_{i,k},X_{i,k})\\ \vdots & \ddots & \vdots & \vdots & \vdots\\ \operatorname{cov}(X_{i,k},X_{i,k}) & \cdots & \operatorname{cov}(X_{i,k},X_{i,k}) & \cdots & \operatorname{cov}(X_{i,k},X_{i,k})\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ \operatorname{cov}(X_{i,k},X_{i,k}) & \cdots & \operatorname{cov}(X_{i,k},X_{i,k}) & \cdots & \operatorname{cov}(X_{i,k},X_{i,k})\end{bmatrix}$$

Recall $\operatorname{cov}(X_{i,k},X_{i,k})=\operatorname{var}(X_{i,k})=1$. As defined in Equation 3.25 we have

$$\operatorname{cov}(\mathbf{X}_k,\mathbf{X}_k)=[1]_{N\times N},\ k\in L_{E,i} \tag{3.31}$$

Corresponding to the vector of the residuals, following Equation 3.19 we have

$$\operatorname{cov}(\boldsymbol{\varepsilon}_i,\boldsymbol{\varepsilon}_i)=\begin{bmatrix}\operatorname{cov}(\varepsilon_i,\varepsilon_i) & \cdots & \operatorname{cov}(\varepsilon_i,\varepsilon_i) & \cdots & \operatorname{cov}(\varepsilon_i,\varepsilon_i)\\ \vdots & \ddots & \vdots & \vdots & \vdots\\ \operatorname{cov}(\varepsilon_i,\varepsilon_i) & \cdots & \operatorname{cov}(\varepsilon_i,\varepsilon_i) & \cdots & \operatorname{cov}(\varepsilon_i,\varepsilon_i)\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ \operatorname{cov}(\varepsilon_i,\varepsilon_i) & \cdots & \operatorname{cov}(\varepsilon_i,\varepsilon_i) & \cdots & \operatorname{cov}(\varepsilon_i,\varepsilon_i)\end{bmatrix}$$

Recall $\operatorname{cov}(\varepsilon_i,\varepsilon_i)=\operatorname{var}(\varepsilon_i)=1$. As defined in Equation 3.25 we have

$$\operatorname{cov}(\boldsymbol{\varepsilon}_i,\boldsymbol{\varepsilon}_i)=[1]_{N\times N} \tag{3.32}$$

Applying Equations 3.25, 3.30, 3.30 and 3.32 and with Equation 3.24, we have

$$\begin{aligned}\operatorname{cov}(\mathbf{p},\mathbf{p})=&\sum_{k\in L_X}\boldsymbol{\lambda}_k\cdot[1]_{N\times N}\cdot\boldsymbol{\lambda}_k^T+\sum_{k\in L_C}\boldsymbol{\lambda}_k\cdot r(C_k)\cdot\boldsymbol{\lambda}_k^T\\ &+\sum_{i=1}^{N}\sum_{k\in L_{E,i}}\boldsymbol{\lambda}_k\cdot[1]_{N\times N}\cdot\boldsymbol{\lambda}_k^T+\sum_{i=1}^{N}\mathbf{w}_i\cdot[1]_{N\times N}\cdot\mathbf{w}_i^T\end{aligned} \tag{3.33}$$

We define the diagonal matrix containing the variance of the input parameters as

$$\boldsymbol{\sigma}_{\mathbf{p}}^{2} = \begin{bmatrix} \mathrm{var}(p_1) & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & \mathrm{var}(p_i) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & \mathrm{var}(p_N) \end{bmatrix} \tag{3.34}$$

Accordingly we define

$$\boldsymbol{\sigma}_{\mathbf{p}} = \begin{bmatrix} \sqrt{\mathrm{var}(p_1)} & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & \sqrt{\mathrm{var}(p_i)} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & \sqrt{\mathrm{var}(p_N)} \end{bmatrix} \tag{3.35}$$

and

$$\boldsymbol{\sigma}_{\mathbf{p}}^{-1} = \begin{bmatrix} \dfrac{1}{\sqrt{\mathrm{var}(p_1)}} & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & \dfrac{1}{\sqrt{\mathrm{var}(p_i)}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & \dfrac{1}{\sqrt{\mathrm{var}(p_N)}} \end{bmatrix} \tag{3.36}$$

From Equations 3.34 - 3.36, we have

$$\boldsymbol{\sigma}_{\mathbf{p}} \cdot \boldsymbol{\sigma}_{\mathbf{p}} = \boldsymbol{\sigma}_{\mathbf{p}}^{2} \tag{3.37}$$

$$\boldsymbol{\sigma}_{\mathbf{p}}^{-1} \cdot \boldsymbol{\sigma}_{\mathbf{p}} = \mathbf{I}_N \tag{3.38}$$

where $\mathbf{I}_N$ stands for the $N \times N$ dimensional identity matrix.

By Equations 3.12, 3.67 and 3.36, we have

$$r(\mathbf{p}) = \boldsymbol{\sigma}_{\mathbf{p}}^{-1} \cdot \mathrm{cov}(\mathbf{p},\mathbf{p}) \cdot \boldsymbol{\sigma}_{\mathbf{p}}^{-1} \tag{3.39}$$

By applying Equation 3.33 with Equation 3.39, we have

$$r(\mathbf{p}) = \sum_{k \in L_X} \left(\boldsymbol{\sigma}_{\mathbf{p}}^{-1} \boldsymbol{\lambda}_k\right) \cdot [1]_{N \times N} \cdot \left(\boldsymbol{\lambda}_k^T \boldsymbol{\sigma}_{\mathbf{p}}^{-1}\right) + \sum_{k \in L_C} \left(\boldsymbol{\sigma}_{\mathbf{p}}^{-1} \boldsymbol{\lambda}_k\right) \cdot r(C_k) \cdot \left(\boldsymbol{\lambda}_k^T \boldsymbol{\sigma}_{\mathbf{p}}^{-1}\right)$$

$$+ \sum_{i=1}^{N} \sum_{k \in L_{E,i}} \left(\boldsymbol{\sigma}_{\mathbf{p}}^{-1} \boldsymbol{\lambda}_k\right) \cdot [1]_{N \times N} \cdot \left(\boldsymbol{\lambda}_k^T \boldsymbol{\sigma}_{\mathbf{p}}^{-1}\right) + \sum_{i=1}^{N} \left(\boldsymbol{\sigma}_{\mathbf{p}}^{-1} \mathbf{w}_i\right) \cdot [1]_{N \times N} \cdot \left(\mathbf{w}_i^T \boldsymbol{\sigma}_{\mathbf{p}}^{-1}\right) \tag{3.40}$$

By Equation 3.36, we have

$$\boldsymbol{\sigma}_{\mathbf{p}}^{-1} = \left(\boldsymbol{\sigma}_{\mathbf{p}}^{-1}\right)^T \tag{3.41}$$

By Equations 3.14, 3.16 and 3.18, the matrix of the loadings associated all the types of the vectors of factors are diagonal. Consequently we have

$$\boldsymbol{\lambda}_k = \boldsymbol{\lambda}_k^T \tag{3.42}$$

From Equations 3.41 and 3.42, we have

$$\boldsymbol{\lambda}_k^T \boldsymbol{\sigma}_{\mathbf{p}}^{-1} = \boldsymbol{\lambda}_k^T \left(\boldsymbol{\sigma}_{\mathbf{p}}^{-1}\right)^T = \left(\boldsymbol{\sigma}_{\mathbf{p}}^{-1} \boldsymbol{\lambda}_k\right)^T \tag{3.43}$$

By Equation 3.20, we have also

$$\mathbf{w}_i = \mathbf{w}_i^T \tag{3.44}$$

From Equations 3.41 and 3.44, we have

$$\mathbf{w}_i^T \boldsymbol{\sigma}_\mathbf{p}^{-1} = \mathbf{w}_i^T \left(\boldsymbol{\sigma}_\mathbf{p}^{-1}\right)^T = \left(\boldsymbol{\sigma}_\mathbf{p}^{-1} \mathbf{w}_i\right)^T \tag{3.45}$$

Applying Equations 3.43 and 3.45 with Equation 3.40, we have

$$\begin{aligned} r(\mathbf{p}) = & \sum_{k \in L_X} \left(\boldsymbol{\sigma}_\mathbf{p}^{-1} \boldsymbol{\lambda}_k\right) \cdot [1]_{N \times N} \cdot \left(\boldsymbol{\sigma}_\mathbf{p}^{-1} \boldsymbol{\lambda}_k\right)^T + \sum_{k \in L_C} \left(\boldsymbol{\sigma}_\mathbf{p}^{-1} \boldsymbol{\lambda}_k\right) \cdot r(C_k) \cdot \left(\boldsymbol{\sigma}_\mathbf{p}^{-1} \boldsymbol{\lambda}_k\right)^T \\ & + \sum_{i=1}^{N} \sum_{k \in L_{E,i}} \left(\boldsymbol{\sigma}_\mathbf{p}^{-1} \boldsymbol{\lambda}_k\right) \cdot [1]_{N \times N} \cdot \left(\boldsymbol{\sigma}_\mathbf{p}^{-1} \boldsymbol{\lambda}_k\right)^T + \sum_{i=1}^{N} \left(\boldsymbol{\sigma}_\mathbf{p}^{-1} \mathbf{w}_i\right) \cdot [1]_{N \times N} \cdot \left(\boldsymbol{\sigma}_\mathbf{p}^{-1} \mathbf{w}_i\right)^T \end{aligned} \tag{3.46}$$

From Equation 3.46, we can decompose $r(\mathbf{p})$ into the correlation matrices associated with the common factors, the factor classes, the exclusive factors and the residuals. When $\mathbf{X}_k$ is associated with a common factor X_k, i.e. $k \in L_X$, we define

$$r(\mathbf{p})\Big|_{X_k} = \left(\boldsymbol{\sigma}_\mathbf{p}^{-1} \boldsymbol{\lambda}_k\right) \cdot [1]_{N \times N} \cdot \left(\boldsymbol{\sigma}_\mathbf{p}^{-1} \boldsymbol{\lambda}_k\right)^T \tag{3.47}$$

When $\mathbf{X}_k$ is associated with a factor class C_k, i.e. $k \in L_C$, we define

$$r(\mathbf{p})\Big|_{C_k} = \left(\boldsymbol{\sigma}_\mathbf{p}^{-1} \boldsymbol{\lambda}_k\right) \cdot r(C_k) \cdot \left(\boldsymbol{\sigma}_\mathbf{p}^{-1} \boldsymbol{\lambda}_k\right)^T \tag{3.48}$$

When $\mathbf{X}_k$ is associated with an exclusive factor $X_{i,k}$, i.e. $k \in L_{E,i}$, we define

$$r(\mathbf{p})\Big|_{X_{i,k}} = \left(\boldsymbol{\sigma}_\mathbf{p}^{-1} \boldsymbol{\lambda}_k\right) \cdot [1]_{N \times N} \cdot \left(\boldsymbol{\sigma}_\mathbf{p}^{-1} \boldsymbol{\lambda}_k\right)^T \tag{3.49}$$

According the vector of the residual $\boldsymbol{\varepsilon}_i$, we define

$$r(\mathbf{p})\big|_{\varepsilon_i} = \left(\boldsymbol{\sigma}_{\mathbf{p}}^{-1}\mathbf{w}_i\right)\cdot[1]_{N\times N}\cdot\left(\boldsymbol{\sigma}_{\mathbf{p}}^{-1}\mathbf{w}_i\right)^T \tag{3.50}$$

Using the notations established in Equations 3.47-3.50, we express Equation 3.40 as

$$r(\mathbf{p}) = \sum_{k\in L_X} r(\mathbf{p})\big|_{X_k} + \sum_{k\in L_C} r(\mathbf{p})\big|_{C_k} + \sum_{i=1}^{N}\sum_{k\in L_{E,i}} r(\mathbf{p})\big|_{X_{i,k}} + \sum_{i=1}^{N} r(\mathbf{p})\big|_{\varepsilon_i} \tag{3.51}$$

We call $r(\mathbf{p})\big|_{X_k}$, $r(\mathbf{p})\big|_{C_k}$, $r(\mathbf{p})\big|_{X_{i,k}}$ and $r(\mathbf{p})\big|_{\varepsilon_i}$ the correlation matrix of $\mathbf{p}$ associated with the common factor, the factor class, the exclusive factor and the residual respectively. We are going to show how these correlation matrices can be defined by the elicitation of expert judgement. We will also prove that these correlation matrices are positive semi-definite.

3.6.3 Correlation Matrix Associated with a Common Factor

The correlation matrix of $\mathbf{p}$ associated with a common factor X_k is defined in Equation 3.47. From Equations 3.14 and 3.36, we have

$$\boldsymbol{\sigma}_{\mathbf{p}}^{-1}\boldsymbol{\lambda}_k = \begin{bmatrix} \dfrac{\lambda_{1,k}}{\sqrt{\mathrm{var}(p_1)}} & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & \dfrac{\lambda_{i,k}}{\sqrt{\mathrm{var}(p_i)}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & \dfrac{\lambda_{N,k}}{\sqrt{\mathrm{var}(p_N)}} \end{bmatrix} \tag{3.52}$$

From Equation 3.6, we have

$$\frac{\lambda_{i,k}}{\sqrt{\mathrm{var}(p_i)}} = r(p_i, X_k) \tag{3.53}$$

where $r(p_i, X_k)$ represents the correlation between p_i and X_k.

Applying Equation 3.53 with Equation 3.52, we have

$$\begin{aligned}\boldsymbol{\sigma}_{\mathbf{p}}^{-1}\boldsymbol{\lambda}_k &= \begin{bmatrix} r(p_1, X_k) & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & r(p_i, X_k) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & r(p_N, X_k) \end{bmatrix} \\ &\equiv r(\mathbf{p}, X_k)\end{aligned} \tag{3.54}$$

The diagonal matrix $r(\mathbf{p}, X_k)$ is called the correlation matrix between the input parameter vector $\mathbf{p}$ and the common factor X_k.

Applying Equations 3.54 with Equation 3.47, we have

$$r(\mathbf{p})\big|_{X_k} = r(\mathbf{p}, X_k) \cdot [1]_{N\times N} \cdot (r(\mathbf{p}, X_k))^T \tag{3.55}$$

By analogy with $h_{i,j}^k$ defined in Equation 3.8, $r(\mathbf{p})\big|_{X_k}$ represents the correlation matrix of $\mathbf{p}$ driven by the common factor X_k. We are going to prove that $r(\mathbf{p})\big|_{X_k}$ is positive semi-define. For this purpose, we define the square-root of $[1]_{N\times N}$ as

$$([1]_{N\times N})^{1/2} = \sqrt{\frac{1}{N}} \cdot [1]_{N\times N} \equiv \left[\sqrt{\frac{1}{N}}\right]_{N\times N} \tag{3.56}$$

Therefore we have

$$\left(\left[\frac{1}{\sqrt{N}}\right]_{N\times N}\right)^{1/2} = \left(\left(\left[\frac{1}{\sqrt{N}}\right]_{N\times N}\right)^{1/2}\right)^T \tag{3.57}$$

and

$$[1]_{N\times N} = \left(\left[\frac{1}{\sqrt{N}}\right]_{N\times N}\right)^{1/2} \cdot \left(\left(\left[\frac{1}{\sqrt{N}}\right]_{N\times N}\right)^{1/2}\right)^{T} \tag{3.58}$$

Applying Equations 3.58 and 3.57 with Equation 3.55, we have

$$r(\mathbf{p})\big|_{X_k} = r(\mathbf{p}, X_k) \cdot \left(\left[\frac{1}{\sqrt{N}}\right]_{N\times N}\right) \cdot \left(\left[\frac{1}{\sqrt{N}}\right]_{N\times N}\right)^{T} \cdot (r(\mathbf{p}, X_k))^{T} \tag{3.59}$$

As defined in Equation 3.54, $r(\mathbf{p}, X_k)$ is a diagonal matrix and therefore we have

$$r(\mathbf{p}, X_k) = (r(\mathbf{p}, X_k))^{T} \tag{3.60}$$

Based on Equations 3.59 and 3.60, we have

$$r(\mathbf{p})\big|_{X_k} = r(\mathbf{p}, X_k) \cdot \left(\left[\frac{1}{\sqrt{N}}\right]_{N\times N}\right) \cdot \left(r(\mathbf{p}, X_k) \cdot \left(\left[\frac{1}{\sqrt{N}}\right]_{N\times N}\right)\right)^{T} \tag{3.61}$$

By Theorem C.2 in Appendix C, the matrix $r(\mathbf{p})\big|_{X_k}$ in Equation 3.61 is positive semi-definite. This property of $r(\mathbf{p})\big|_{X_k}$ is to be used later to prove that $r(\mathbf{p})$ is positive semi-definite.

3.6.4 Correlation Matrix Associated with a Factor Class

The correlation matrix of **p** associated with a factor class C_k is defined in Equation 3.48. From Equations 3.16 and 3.36, we

$$\boldsymbol{\sigma}_{\mathbf{p}}^{-1}\boldsymbol{\lambda}_k = \begin{bmatrix} \frac{\lambda_{1,k}}{\sqrt{\mathrm{var}(p_1)}} & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & \frac{\lambda_{i,k}}{\sqrt{\mathrm{var}(p_i)}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & \frac{\lambda_{N,k}}{\sqrt{\mathrm{var}(p_N)}} \end{bmatrix} \tag{3.62}$$

Applying Equation 3.53 with Equation 3.62, we have

$$\begin{aligned} \boldsymbol{\sigma}_{\mathbf{p}}^{-1}\boldsymbol{\lambda}_k &= \begin{bmatrix} r(p_1, X_{1,k}) & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & r(p_i, X_{i,k}) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & r(p_N, X_{N,k}) \end{bmatrix} \\ &\equiv r(\mathbf{p}, C_k) \end{aligned} \tag{3.63}$$

The diagonal matrix $r(\mathbf{p}, C_k)$ is called the correlation matrix between the input parameter vector $\mathbf{p}$ and the factor class C_k. This matrix needs to be built through elicitation of expert judgement.

Applying Equation 3.62 with Equation 3.48, we have

$$r(\mathbf{p})\big|_{C_k} = r(\mathbf{p}, C_k) \cdot r(C_k) \cdot (r(\mathbf{p}, C_k))^T \tag{3.64}$$

By Theorem C.3 in Appendix C, $r(\mathbf{p})\big|_{C_k}$ defined in Equation 3.64 is semi-definite when $r(C_k)$ in Equation 3.20 is kept positive semi-definite.

As discussed in Section 3.4, a factor class in practice is kept "thin" by agglomerating the included factors into a few factors each of which is common to a subset of the input parameters. For a thin factor class, it is feasible to build a positive semi-definite

$r(C_k)$ through elicitation of expert judgement [Ghosh and Henderson 2003]. The LC users in the example in Section 3.4, for example, are agglomerated into the users of UWC, denoted by $C_k^{(U)}$, and the users of non-UWC, denoted by $C_k^{(N)}$. Five of the eight input parameters are associated with $C_k^{(U)}$ and the other three are associated with $C_k^{(N)}$. The associated $r(C_k)$ can then be defined as

$$r(C_k)=\left[\begin{array}{ccccc:ccc} 1 & 1 & 1 & 1 & 1 & & & \\ 1 & 1 & 1 & 1 & 1 & & & \\ 1 & 1 & 1 & 1 & 1 & \multicolumn{3}{c}{r\left(C_k^{(U)},C_k^{(N)}\right)} \\ 1 & 1 & 1 & 1 & 1 & & & \\ 1 & 1 & 1 & 1 & 1 & & & \\ \hdashline \multicolumn{5}{c:}{} & 1 & 1 & 1 \\ \multicolumn{5}{c:}{r\left(C_k^{(U)},C_k^{(N)}\right)} & 1 & 1 & 1 \\ \multicolumn{5}{c:}{} & 1 & 1 & 1 \end{array}\right]$$

Therefore $r\left(C_k^{(U)},C_k^{(N)}\right)$ is the only one that needs to be elicited. The outcome $r(C_k)$ is automatically positive definite.

3.6.5 Correlation Matrix Associated with an Exclusive Factor

The correlation matrix of **p** associated with an exclusive factor is defined in Equation 3.49. From Equations 3.18 and 3.36, we

$$\boldsymbol{\sigma}_{\mathbf{p}}^{-1}\boldsymbol{\lambda}_k=\begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & \dfrac{\lambda_{i,k}}{\sqrt{\mathrm{var}(p_i)}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \\ & & & & 0 \end{bmatrix} \tag{3.65}$$

Applying Equation 3.53 with Equation 3.65, we have

$$\boldsymbol{\sigma}_{\mathbf{p}}^{-1}\boldsymbol{\lambda}_k = \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & r(p_i, X_{i,k}) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \\ & & & & 0 \end{bmatrix} \equiv r(\mathbf{p}, X_{i,k}) \tag{3.66}$$

Corresponding to the matrix of loadings $\boldsymbol{\lambda}_k$, $k \in L_{E,i}$ as defined in Equation 4.N6, the matrix $r(\mathbf{p}, X_{i,k})$ contains only one non-zero $r(p_i, X_{i,k})$ that needs to be elicited from the experts.

Applying Equation 3.66 with Equation 3.49, we have

$$r(\mathbf{p})\Big|_{X_{i,k}} = r(\mathbf{p}, X_{i,k}) \cdot [1]_{N\times N} \cdot (r(\mathbf{p}, X_{i,k}))^T$$

By the calculations of the matrices we have

$$r(\mathbf{p})\Big|_{X_{i,k}} = \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & r^2(p_i, X_{i,k}) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} \tag{3.67}$$

The $r(\mathbf{p})\Big|_{X_{i,k}}$ has 0 for all the cells except the i^{th} diagonal one. Apparently $r(\mathbf{p})\Big|_{X_{i,k}}$ is associated only with $r(p_i, p_i)$ in $r(\mathbf{p})$. It is consistent with the definition that the exclusive factor impacts only the associated input parameter and therefore does not contribute to the correlation among the parameters. Because of $r^2(p_i, X_{i,k}) \geq 0$, $r(\mathbf{p})\Big|_{X_{i,k}}$ is positive semi-definite by Theorem C.1 in Appendix C.

3.6.6 Correlation Matrix Associated with the Residual

The correlation matrix of **p** associated with a residual is defined in Equation 3.50. From Equations 3.20 and 3.36, we

$$\boldsymbol{\sigma}_{\mathbf{p}}^{-1}\mathbf{w}_i = \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & \dfrac{w_i}{\sqrt{\mathrm{var}(p_i)}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} \tag{3.68}$$

Therefore the matrix $\boldsymbol{\sigma}_{\mathbf{p}}^{-1}\mathbf{w}_i$ contains only one non-zero cell on the diagonal corresponding to the associated residual ε_i. Different from all the factors, the matrix $\boldsymbol{\sigma}_{\mathbf{p}}^{-1}\mathbf{w}_i$ is not to be derived based on the elicitation regarding the associated residual ε_i. Rather, we will derive $\dfrac{w_i}{\sqrt{\mathrm{var}(p_i)}}$ from the elicitation regarding the factors.

Recalling Equation 3.2, $\mathrm{var}(p_i) = \sum_k \lambda_{i,k}^2 + w_i^2$, and Equation 3.6 $\lambda_{i,k} = r(p_i, X_{i,k})\sqrt{\mathrm{var}(p_i)}$. Applying Equation 3.6 with Equation 3.2, we have

$$\mathrm{var}(p_i) = \mathrm{var}(p_i)\left(\sum_{k\in L_X} r^2(p_i, X_k) + \sum_{k\in L_C} r^2(p_i, X_{i,k}) + \sum_{k\in L_{E,i}} r^2(p_i, X_{i,k}) \right) + w_i^2 \tag{3.69}$$

We then have

$$1 = \sum_{k\in L_X} r^2(p_i, X_k) + \sum_{k\in L_C} r^2(p_i, X_{i,k}) + \sum_{k\in L_{E,i}} r^2(p_i, X_{i,k}) + \frac{w_i^2}{\mathrm{var}(p_i)} \tag{3.70}$$

The four terms on the right side of Equation 3.70 represent the four parts of $r(p_i, p_i)$ corresponding in the sequence to: (a) X_k the common factors X_k, $k \in L_X$; (b) the

correlated factors $X_{i,k}$, $k \in L_C$; (c) the exclusive factors $X_{i,k}$, $k \in L_{E,i}$; and (d) the residual uncertainty ε_i. Parts (a), (b) and (c) can be defined by the elicitation of the correlations between the input parameter and the associated factors. From Equation 3.70, we define

$$q_i = \sum_{k \in L_X} r^2(p_i, X_k) + \sum_{k \in L_C} r^2(p_i, X_{i,k}) + \sum_{k \in L_{E,i}} r^2(p_i, X_{i,k}) \tag{3.71}$$

Because $\frac{w_i^2}{\text{var}(p_i)} \geq 0$, by Equations 3.70 and 3.71, we must have

$$0 \leq q_i \leq 1 \tag{3.72}$$

Applying Equation 3.71 with Equation 3.70, we have

$$\frac{w_i^2}{\text{var}(p_i)} = 1 - q_i \tag{3.73}$$

We can see that q_i stands for the proportion of $\text{var}(p_i)$ that is explained by all the factors associated with p_i. Inequality 3.72 offers a constraint to monitor the experts' assessments during the elicitation. With q_i derived by Equation 3.71 through elicitation, applying Equation 3.73 with Equation 3.68 we have

$$\boldsymbol{\sigma}_{\mathbf{p}}^{-1}\mathbf{w}_i = \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & \sqrt{1-q_i} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & \begin{matrix} 0 \\ 0 \end{matrix} \end{bmatrix} \tag{3.74}$$

Applying Equation 3.74 with Equation 3.50, we have

$$r(\mathbf{p})\big|_{\varepsilon_i} = \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & \sqrt{1-q_i} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} \cdot [1]_{N\times N} \cdot \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & \sqrt{1-q_i} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1-q_i & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{bmatrix}$$

(3.75)

By Theorem C.1 in Appendix C, $r(\mathbf{p})\big|_{\varepsilon_i}$ is positive semi-definite when $0 \le q_i \le 1$ is satisfied. For all the residuals ε_i, $i = 1, \cdots, N$, a correlation matrix can be defined in a compact form as

$$r(\mathbf{p})\big|_{\varepsilon} = \sum_{i=1,N} r(\mathbf{p})\big|_{\varepsilon_i} \tag{3.76}$$

Applying Equation 3.75 with 3.76, we have

$$r(\mathbf{p})\big|_{\varepsilon} = \begin{bmatrix} 1-q_1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1-q_i & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 1-q_N \end{bmatrix} \tag{3.77}$$

By Theorem C.4 in Appendix C, the $r(\mathbf{p})\big|_{\varepsilon}$ is positive semi-definite when all $r(\mathbf{p})\big|_{\varepsilon_i}$, $i = 1, \cdots, N$ are positive semi-definite.

3.6.7 Correlation Matrix as Sum of Correlation Matrices Associated with the Various Types of Factors

With the definitions given in Equations 3.55, 3.64, 3.67, 3.75, 3.77 and 3.76, the overall correlation matrix of N input parameters can be derived from Equation 3.51 as

$$r(\mathbf{p}) = \sum_{k \in L_X} r(\mathbf{p})\big|_{X_k} + \sum_{k \in L_C} r(\mathbf{p})\big|_{C_k} + \sum_{i=1,N} \sum_{k \in L_{E,i}} r(\mathbf{p})\big|_{X_{i,k}} + r(\mathbf{p})\big|_{\varepsilon} \tag{3.78}$$

In Equation 3.78, the matrices $r(\mathbf{p})\big|_{X_k}$ and $r(\mathbf{p})\big|_{C_k}$ are the correlation contribution of the common factors and factor classes respectively; $r(\mathbf{p})\big|_{X_{i,k}}$ and $r(\mathbf{p})\big|_{\varepsilon}$ are matrix associated with the exclusive factors and the residual terms respectively. As developed above, these matrices can be built based on the elicitation of the correlations between the input parameters and the associated factors and the correlations among the factors in a class. Following the above procedure, these correlation matrices can be kept positive semi-definite. As a result the correlation matrix $r(\mathbf{p})$ can be kept positive semi-definite by Theorem C.4 in Appendix C. When $q_i < 1$ for some $i = 1, \cdots, N$, then $r(\mathbf{p})\big|_{\varepsilon}$ is positive definite and as a result $r(\mathbf{p})$ is positive definite by Theorem C.4 in Appendix C.

From Equation 3.78, all exclusive factors associated with one input parameter can be represented by a single one. From operational perspective, however, we don not do that way. Elicitation starts with identifying all the major factors for each input parameter. At the second step, factors are then classified. Therefore, we don't know if a factor belongs to the exclusive category or not until we identify all the major factors. Furthermore, with all the major factors identified, experts are more confident with assessing the associated correlations.

3.7 Definition of an Input Parameter Family

3.7.1 Motivation

The primary motivation for establishing the parameter family concept is to reduce the elicitation workload regarding the variance. With the correlation matrix of a set of input parameters built, we need the variance of input parameter to build the covariance matrix. It can be done through elicitation of the expert judgement [Cooke 1991; Garthwaite, Kadane et al. 2005; Kurowicka and Cooke 2006; O'Hagan, Buck et al. 2006]. For a large set of input parameters, it requires much elicitation workload however. We therefore need to develop a way to reduce the workload. A group of input parameters can be defined as physically similar events. For example, a group of RSSB-SRM precursors are defined as *road vehicle (RV) driver error causing RV struck by train on level crossing L,* where *L* denotes one of the eight types of level crossings (LCs) [RSSB 2004]. We can see that this group of the precursors are defined as the same event associated with different types of level crossings. We are going to define such a group of input parameters to be a family. When input parameters are occurrence rate of a Poisson process, their sum is the occurrence rate of the family. The idea is to elicit the variance in the occurrence rate of the family and derive the variance of the precursors in the family. By eliciting on the family instead of individual parameters, the elicitation workload is reduced significantly.

The second motivation is to facilitate the assessment. Experts are more confident in giving their assessment on higher level events, as shown in the case of RSSB-SRM. By eliciting on the family level rather than individual parameters, we can make the elicitation easier to experts.

3.7.2 Definition

We define a group of input parameters to be a family, when: (1) the input parameters have physically similar interpretation for which a basic event can be defined; (2) a set of differences can be identified among the input parameters; (3) affected by some differences, each individual parameter derives from the basic event. Therefore, after the effect of the differences is accounted, the input parameters can be treated as the

basic event. The family members therefore share a common part of uncertainty due to the basic event. The factors associated with the basic event are invariant factors; the factors associated with variations are variant factors.

We are going to develop the mathematical definition of the family based on the uncertainty model as defined in Equation 3.4. We denote a family by f. Suppose two input parameters p_i and p_j belong to the family f, denoted by $p_i, p_j \in f$. Recall that the uncertainty in p_i and p_j can be defined on their factors of three categories as in Equation 3.4 as

$$p_i - \mu_i = \sum_{k \in L_X} \lambda_{i,k} X_k + \sum_{k \in L_C} \lambda_{i,k} X_{i,k} + \sum_{k \in L_{E,i}} \lambda_{i,k} X_{i,k} + w_i \varepsilon_i$$

$$p_j - \mu_j = \sum_{k \in L_X} \lambda_{j,k} X_k + \sum_{k \in L_C} \lambda_{j,k} X_{j,k} + \sum_{k \in L_{E,j}} \lambda_{j,k} X_{j,k} + w_j \varepsilon_j$$

We define the common factors that are associated with the basic event as the invariant factors of the family, i.e.

$$I(f) = \{X_k \,|\, X_k \text{ is associated with the basic event}\}$$

We define the other factors that are associated with the differences among the family as the variant factors, i.e.

$$V(f) = \{X_k \,|\, X_k \text{ is not associated with the basic event}\} \bigcup \{X_{i,k} \,|\, k \in L_{E,i}\} \bigcup \{X_{i,k} \,|\, k \in L_C\}$$

We can see that the common factors are split into $I(f)$ and $V(f)$ depending on if they are associated with the basic. The uncertainty model of the family can then be defined based on Equation 3.4 as

$$p_i - \mu_i = \sum_{\substack{X_k \in I(f) \\ k \in L_X}} \lambda_{i,k} X_k + \sum_{\substack{X_k \in V(f) \\ k \in L_X}} \lambda_{i,k} X_k + \sum_{k \in L_C} \lambda_{i,k} X_{i,k} + \sum_{k \in L_{E,i}} \lambda_{i,k} X_{i,k} + w_i \varepsilon_i$$

$$p_j - \mu_j = \sum_{\substack{X_k \in I(f) \\ k \in L_X}} \lambda_{j,k} X_k + \sum_{\substack{X_k \in V(f) \\ k \in L_X}} \lambda_{j,k} X_k + \sum_{k \in L_C} \lambda_{j,k} X_{j,k} + \sum_{k \in L_{E,j}} \lambda_{j,k} X_{j,k} + w_j \varepsilon_j \quad (3.79)$$

For some families, the members can also share the same residual term that is not affected by any difference across the family. In this situation, we call the residual term family residual that is an invariant factor. The uncertainty model of the family correspondingly becomes

$$p_i - \mu_i = \sum_{\substack{X_k \in I(f) \\ k \in L_X}} \lambda_{i,k} X_k + \sum_{\substack{X_k \in V(f) \\ k \in L_X}} \lambda_{i,k} X_k + \sum_{k \in L_C} \lambda_{i,k} C_{i,k} + \sum_{k \in L_{E,i}} \lambda_{i,k} X_{i,k} + w_i \varepsilon_f$$

$$p_j - \mu_j = \sum_{\substack{X_k \in I(f) \\ k \in L_X}} \lambda_{j,k} X_k + \sum_{\substack{X_k \in V(f) \\ k \in L_X}} \lambda_{j,k} X_k + \sum_{k \in L_C} \lambda_{j,k} C_{j,k} + \sum_{k \in L_{E,j}} \lambda_{j,k} X_{j,k} + w_j \varepsilon_f \quad (3.80)$$

where ε_f stands for the family residual.

Therefore the uncertainty model of the family can be defined in two situations depending on the property of the residual uncertainty terms in the family.

Based on Equations 3.79 and 3.80, the variance of the family members can be split into the common part due to the invariant factors and the part due to the variant factors as defined as

$$\mathrm{var}(p_i)\big|_{I(f)} = \sum_{\substack{X_k \in I(f) \\ k \in L_X}} \lambda_{i,k}^2 + \delta(\varepsilon_f) \cdot w_i^2 \quad (3.81)$$

$$\mathrm{var}(p_i)\big|_{V(f)} = \sum_{\substack{X_k \in V(f) \\ k \in L_X}} \lambda_{i,k}^2 + \sum_{k \in L_C} \lambda_{i,k}^2 + \sum_{k \in L_{E,i}} \lambda_{i,k}^2 + (1 - \delta(\varepsilon_f)) w_i^2 \quad (3.82)$$

where

$$\delta(\varepsilon_f) = \begin{cases} 1, \text{ when residual is an invariant factor} \\ 0, \text{ otherwise} \end{cases}$$

By Equation 3.2 we have

$$\text{var}(p_i) = \text{var}(p_i)\big|_{I(f)} + \text{var}(p_i)\big|_{v(f)} \tag{3.83}$$

We call $\frac{p_i}{\mu_i}$ normalized input parameter, where $\mu_i = E(p_i)$. For $p_i, p_j \in f$, $\text{var}\left(\frac{p_i}{\mu_i}\right)\Big|_{I(f)}$ and $\text{var}\left(\frac{p_i}{\mu_i}\right)\Big|_{I(f)}$ are associated with the basic event that is the same for the family. Therefore we have

$$\text{var}\left(\frac{p_i}{\mu_i}\right)\Big|_{I(f)} = \text{var}\left(\frac{p_j}{\mu_j}\right)\Big|_{I(f)}, \tag{3.84}$$

From Equations 3.84, we have

$$\frac{\text{var}(p_i)\big|_{I(f)}}{\text{var}(p_j)\big|_{I(f)}} = \frac{\mu_i^2}{\mu_j^2} \tag{3.85}$$

When the input parameters are occurrence rate, the sum of the family members is the occurrence rate of the family. The variance of the family occurrence rate, called the variance of the family, is then meaningful and can be derived by eliciting the percentile of the waiting time [Garthwaite, Kadane et al. 2005]. The variance of the family members can then be derived from the family variance.

Once a whole set of input parameters are partitioned into families, the covariance matrix can be partitioned into blocks correspondingly as to be discussed next.

3.7.3 Covariance Matrix Partitioned by Families

Suppose that there is a group of parameters denoted by $\mathbf{p} = [p_1, \cdots, p_n, \cdots p_N]^T$. The parameters are partitioned into Z families denoted by f_z, $z = 1, \cdots, Z$. Accordingly, the overall correlation matrix $r(\mathbf{p})$ is partitioned into blocks as

$$r(\mathbf{p}) = \begin{bmatrix} r(f_1) & \cdots & r(f_1, f_z) & \cdots & r(f_1, f_y) & \cdots & r(f_1, f_Z) \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ r(f_z, f_1) & \cdots & r(f_z) & \cdots & r(f_z, f_y) & \cdots & r(f_z, f_Z) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ r(f_y, f_1) & \cdots & r(f_y, f_z) & \cdots & r(f_y) & \cdots & r(f_y, f_Z) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ r(f_Z, f_1) & \cdots & r(f_Z, f_z) & \cdots & r(f_Z, f_y) & \cdots & r(f_Z) \end{bmatrix}$$

where $r(f_z)$ stands for the correlation matrix of f_z; $r(f_z, f_y)$ stands for the correlation block between families f_z and f_y.

The diagonal blocks represent correlations within families; the off-diagonal blocks represent the correlations between families. The methods developed previously can be used to build the correlation matrix.

The variance of the parameters $\text{var}(p_i)$ will be derived within the associated family f_z. The method is going to be developed next. Given $\text{var}(p_i)$, $i = 1, \cdots, N$, recall the definition in Equation 3.35,

$$\boldsymbol{\Lambda}^{\frac{1}{2}}(\mathbf{p}) = \begin{bmatrix} \sqrt{\mathrm{var}(p_1)} & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & \sqrt{\mathrm{var}(p_n)} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & \sqrt{\mathrm{var}(p_N)} \end{bmatrix}$$

Consequently the covariance matrix is

$$\mathrm{cov}(\mathbf{p}) = \boldsymbol{\Lambda}^{\frac{1}{2}}(\mathbf{p}) \cdot r(\mathbf{p}) \cdot \boldsymbol{\Lambda}^{\frac{1}{2}}(\mathbf{p})$$

Because $r(\mathbf{p})$ can be guaranteed to be positive semi-definite, $\mathrm{cov}(\mathbf{p})$ is positive semi-definite by Theorem C.3 in Appendix C.

In the next section, the correlation matrix of an input parameter family is to be developed based on Equations 3.79 and 3.80 respectively. After that, we continue to develop the method for eliciting the variance of input parameters within the family.

3.8 Correlation Matrix of a Parameter Family

3.8.1 Family with Independent Residuals

When the family members have independent residuals ε_i, the linear model is defined as in Equation 3.79. This model is the same as defined in Equation 3.4 except that the common factors are split into two parts for invariant factors and variant factors respectively. Correspondingly the correlation matrix of the family f therefore can be adapted from Equation 3.78 as

$$r(f) = \sum_{\substack{X_k \in I(f) \\ k \in L_X}} r(f)\Big|_{X_k} + \sum_{\substack{X_k \in V(f) \\ k \in L_X}} r(f)\Big|_{X_k} + \sum_{k \in L_C} r(f)\Big|_{C_k} + \sum_{i=1,N} \sum_{k \in L_{E,i}} r(f)\Big|_{X_{i,k}} + r(f)\Big|_{\varepsilon} \tag{3.86}$$

where $r(f)$ stands for the correlation matrix of f; $r(f)\big|_{X_k}$, $r(f)\big|_{C_k}$, $r(f)\big|_{X_{i,k}}$ and $r(f)\big|_{\varepsilon}$ are the contributions from the common factors, factor classes, exclusive factors and the residuals as defined in Equations 3.55, 3.64, 3.67 and 3.76 respectively.

3.8.2 Family Residual as an Invariant Factor

When the family has the family residual as an invariant factor, the linear model is defined as in Equation 3.80. The family's correlation matrix can be derived from Equation 3.86 as

$$r(f) = \sum_{\substack{X_k \in I(f) \\ k \in L_X}} r(f)\big|_{X_k} + \sum_{\substack{X_k \in V(f) \\ k \in L_X}} r(f)\big|_{X_k} + \sum_{k \in L_C} r(f)\big|_{C_k} + \sum_{i=1,N} \sum_{k \in L_{E,i}} r(f)\big|_{X_{i,k}} + r(f)\big|_{\varepsilon_f} \tag{3.87}$$

where $r(f)\big|_{\varepsilon_f}$ stands for the correlation matrix associated with the family residual ε_f.

We can see that the only difference between Equations 3.86 and 3.87 is the replacement of $r(f)\big|_{\varepsilon}$ with $r(f)\big|_{\varepsilon_f}$. Because ε_f is a common factor, we are going to derive the matrix $r(f)\big|_{\varepsilon_f}$ based on Equation 3.55.

Based on Equation 3.73, we have

$$r(p_i, \varepsilon_f) = \sqrt{\frac{w_i^2}{\mathrm{var}(p_i)}} = \sqrt{1 - q_i}$$

where q_i, as defined in Equation 3.71, is the proportion of $\mathrm{var}(p_i)$ that is explained by the factors.

Based on Equation 3.54, we define the correlation matrix between $\mathbf{p}$ and ε_f as

$$r(\mathbf{p},\varepsilon_f)=\begin{bmatrix} r(p_1,\varepsilon_f) & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & r(p_i,\varepsilon_f) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & r(p_m,\varepsilon_f) \end{bmatrix}$$

$$=\begin{bmatrix} \sqrt{1-q_1} & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & \sqrt{1-q_i} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & \sqrt{1-q_m} \end{bmatrix}$$

(3.88)

where *m* stands for the size of f, i.e. the number of the input parameters in f.

By Equation 3.55, the correlation matrix of the family associated with ε_f can then derived as

$$r(f)\big|_{\varepsilon_f} = r(\mathbf{p},\varepsilon_f)\cdot[1]_{m\times m}\cdot r(\mathbf{p},\varepsilon_f) \tag{3.89}$$

The correlation matrix of f with the family residual ε_f can then be defined by Equation 3.87 with $r(f)\big|_{\varepsilon}$ replaced with $r(f)\big|_{\varepsilon_f}$. As a summary, the process for building the family correlation matrix is demonstrated in Fig. 3.3.

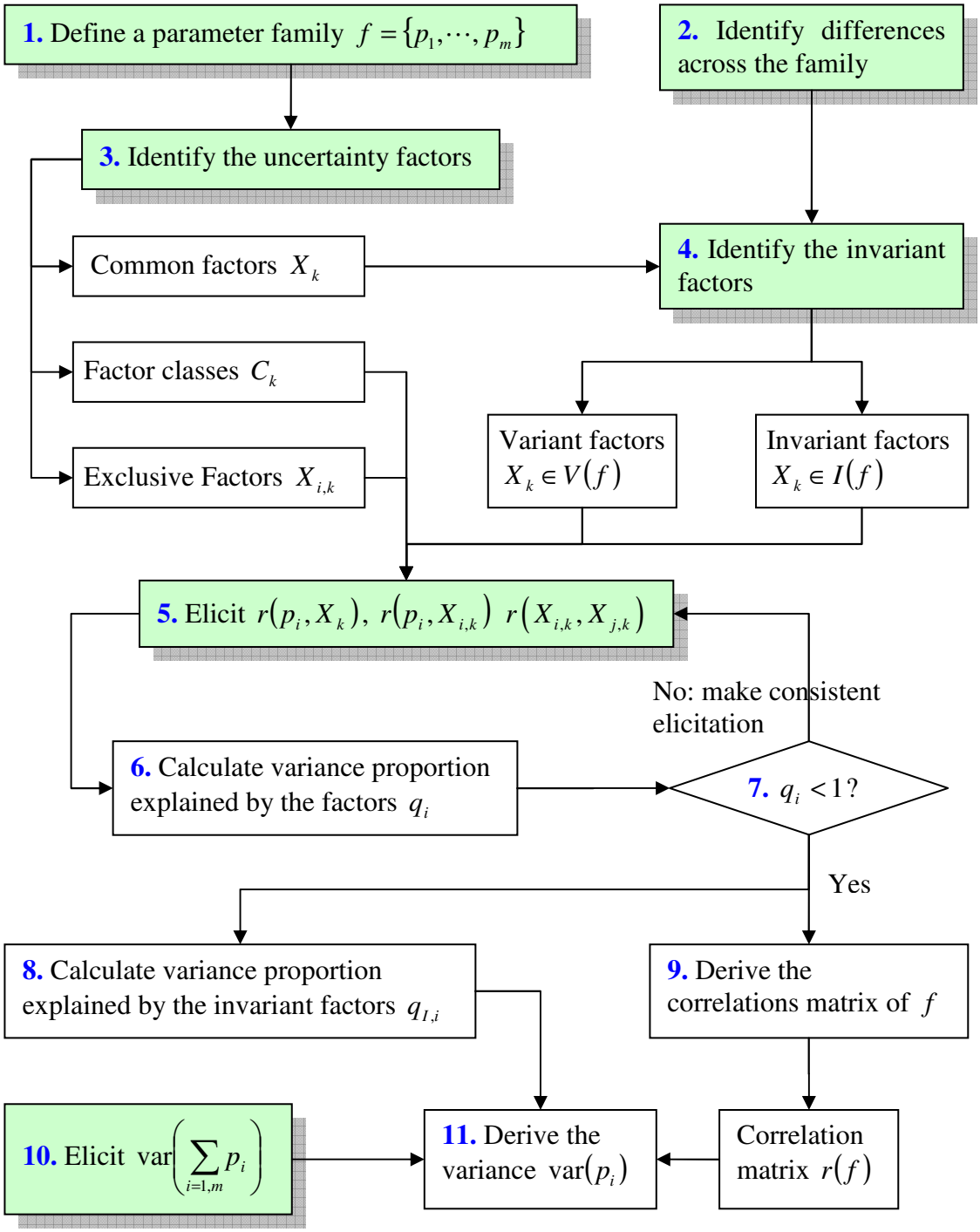


Fig. 3.3 Flow chart for building the family's correlation matrix and the variance of parameters in the family through elicitation

3.9 Deriving Variance of Parameters in a Family

In this section, we will develop the method for deriving variance of family members based on elicitation of the family variance.

By applying Equation 3.6 with Equations 3.81 and 3.82 we have

$$\mathrm{var}(p_i)\big|_{I(f)} = \mathrm{var}(p_i)\cdot\left(\sum_{\substack{X_k\in I(f)\\ k\in L_X}} r^2(p_i,X_k)+\delta(\varepsilon_f)\cdot(1-q_i)\right) \tag{3.90}$$

$$\begin{aligned}\mathrm{var}(p_i)\big|_{V(f)} = \mathrm{var}(p_i)&\left(\sum_{\substack{X_k\in V(f)\\ k\in L_X}} r^2(p_i,X_k)+\sum_{k\in L_C} r^2(p_i,C_{i,k})+\sum_{k\in L_{E,i}} r^2(p_i,X_{i,k})\right)\\ &+\mathrm{var}(p_i)\cdot(1-\delta(\varepsilon_f))\cdot(1-q_i)\end{aligned} \tag{3.91}$$

where q_i is defined as in Equation 3.71.

Based on Equations 3.90 and 3.91, we define

$$q_{I,i} = \sum_{\substack{X_k\in I(f)\\ k\in L_X}} r^2(p_i,X_k)+\delta(\varepsilon_f)\cdot(1-q_i) \tag{3.92}$$

$$q_{V,i} = \sum_{\substack{X_k\in V(f)\\ k\in L_X}} r^2(p_i,X_k)+\sum_{k\in L_C} r^2(p_i,C_{i,k})+\sum_{k\in L_{E,i}} r^2(p_i,X_{i,k})+(1-\delta(\varepsilon_f))\cdot(1-q_i) \tag{3.93}$$

Analogous to q_i, $q_{I,i}$ and $q_{V,i}$ stand for the proportion of $\mathrm{var}(p_i)$ explained by the invariant factors and the variant factors respectively. Based on Equation 3.83 we have

$$q_{I,i} + q_{V,i} = 1 \tag{3.94}$$

For $p_i, p_j \in f$, from Equations 3.90 and 3.92 we have

$$\frac{\mathrm{var}(p_i)}{\mathrm{var}(p_j)} = \frac{\mathrm{var}(p_i)\big|_{I(f)}}{\mathrm{var}(p_j)\big|_{I(f)}} \frac{q_{I,j}}{q_{I,i}} \quad . \tag{3.95}$$

By applying Equation 3.85 with Equation 3.95, we have

$$\frac{\mathrm{var}(p_i)}{\mathrm{var}(p_j)} = \frac{q_{I,j}}{q_{I,i}} \frac{\mu_i^2}{\mu_j^2} \quad . \tag{3.96}$$

In Equation 3.96, $q_{I,i}$ and $q_{I,j}$ can be calculated from the elicitation on the factors by Equations 3.92 and 3.71. The means μ_i and μ_j can be elicited from expert judgement [Clemen, Fischer et al. 2000; Garthwaite, Kadane et al. 2005; O'Hagan, Buck et al. 2006]. Therefore the ratio of the variance between two family members can be derived by Equation 3.96 through elicitation. If we select a family member as the reference denoted by p_r, we can define $\mathrm{var}(p_i)$, $p_i \in f$ in terms of $\mathrm{var}(p_r)$ based on Equation 3.96 as

$$\mathrm{var}(p_i) = \frac{\mu_i^2}{\mu_r^2} \frac{q_{I,r}}{q_{I,i}} \mathrm{var}(p_r) \tag{3.97}$$

When the input parameters are occurrence rates, their sum is meaningful as the occurrence rate of the family as defined as

$$p_f = \sum_{p_i \in f} p_i \tag{3.98}$$

From Equation 3.98, we have

$$\mathrm{var}(p_f) = \sum_{p_i \in f} \sum_{p_j \in f} \mathrm{cov}(p_i, p_j) = \sum_{p_i \in f} \sum_{p_j \in f} r(p_i, p_j) \sqrt{\mathrm{var}(p_i) \cdot \mathrm{var}(p_j)} \qquad (3.99)$$

In Equation 3.99, the family variance $\mathrm{var}(p_f)$ is to be elicited. We can then derive $\mathrm{var}(p_i)$ for $p_i \in f$ from $\mathrm{var}(p_f)$.

By applying Equation 3.97 with Equation 3.99, we have

$$\mathrm{var}(p_f) = \left(\sum_i \sum_j r(p_i, p_j) \cdot \frac{\mu_i}{\mu_r} \cdot \frac{\mu_j}{\mu_r} \cdot \sqrt{\frac{q_{I,r}}{q_{I,i}} \cdot \frac{q_{I,r}}{q_{I,j}}} \right) \cdot \mathrm{var}(p_r) \qquad (3.100)$$

From Equation 3.100, $\mathrm{var}(p_r)$ can be solved in terms of $\mathrm{var}(p_f)$. The $\mathrm{var}(p_i)$ for $p_i \in f$ can then calculated by Equation 3.97.

As a summary, the process for building the family correlation matrix and calculating the variance of the family members is shown in Fig. 3.3. We can see a significant reduction of the elicitation workload when we elicit the variance within the family rather for the input parameters separately. Furthermore, the discussion with the experts in RSSB shows that the experts are more confident in giving their assessment on higher level events.

3.10 Qualitative Assessment and Benchmark of Correlations

Elicitation of the correlations between two random variables is the basic input for building the correlation matrix of a group of input parameters based on the methods developed in the previous sections. Six types of methods for eliciting dependence between two random variables have been summarized and discussed in [Clemen, Fischer et al. 2000]. Three desirable characteristics of a good elicitation method are summarized in the paper. Firstly, a good elicitation method should have rigorous foundations that are defensible in terms of probability theory. Secondly a good method should be a general one that can be used in a wide variety of situations.

Thirdly a good method should be easy to implement and be able to be linked directly to the modelling procedure.

Two of the six types of elicitation methods summarized in [Clemen, Fischer et al. 2000] present the correlation with a continuous line. Two numbers are marked on the both ends of the line. The smaller one on the left side represents the weakest correlation, i.e. 0; while the larger one on the right side represents strongest correlation, i.e. 1. The experts are asked to mark on the line a point of which the distance from the both ends represents the belief of the correlation strength between two random variables. A numerical measure of the correlation is then mapped linearly from the mark position. Because the experts actually are asked to compare the correlations between different pairs of random variables rather than making quantitative assessment directly, these two methods can be categorized as qualitative assessment. They are ranked as the easiest way to implement. However, the linear numerical mapping implies a strong assumption and makes the two methods less rigorous. The mapping however can be done in another way as employed in the Excel Add-in Crystal Ball [Clemen, Fischer et al. 2000]. For a given pair of random variables of preset correlation, the scatter plot is made on the samples and presented on the screen. The experts are then asked to draw a position on the continuous line to shown their belief of the correlation strength. By repeating this process many times, a map can be made between the positions and the underlying correlations. By assuming that the experts can perceive the correlations in a way consistent with their subjective correlation perception, we think this map can be used as benchmark in the qualitative assessment of the correlations.

A combination of the qualitative assessment and the benchmark as defined above is both easy to implement and rigorous in term of probability theory. It is therefore believed to be a good method by the standards summarized in [Clemen, Fischer et al. 2000]. We continue to define the procedure in details. Rather that using the continuous line to represent the correlation, we use finite qualitative correlation levels as agreed with the experts for our case studies.

At the first step, the correlations are elicited from the expert in qualitative levels defined as: Very weak (VW), Weak (W), Moderate (M), Strong (S), Very strong (VS). The number of the levels can be made based on the expert's distinguishing ability with reference to the literature [Clemen, Fischer et al. 2000].

At the second step, benchmark is conducted to map the qualitative assessments into numerical intervals. Take a pair of normal random variables defined as

$$(X,Y) \sim N(\mu,\Sigma)$$

where

$$\mu = (0,0)^T$$

$$\Sigma = \begin{bmatrix} 1 & \rho_{XY} \\ \rho_{XY} & 1 \end{bmatrix}$$

So X and Y are standard normal with correlation coefficient ρ_{XY}. By setting ρ_{XY} randomly between 0 and 1, samples of X and Y can be generated. As demonstrated in Fig. 3.4, the scatter plot is then made on the samples and presented on the screen. The expert is then asked to assess the correlation in terms of the same qualitative levels as used in the elicitation at the first step. The underlying correlation coefficient and the expert's qualitative assessment are recoded for benchmark.

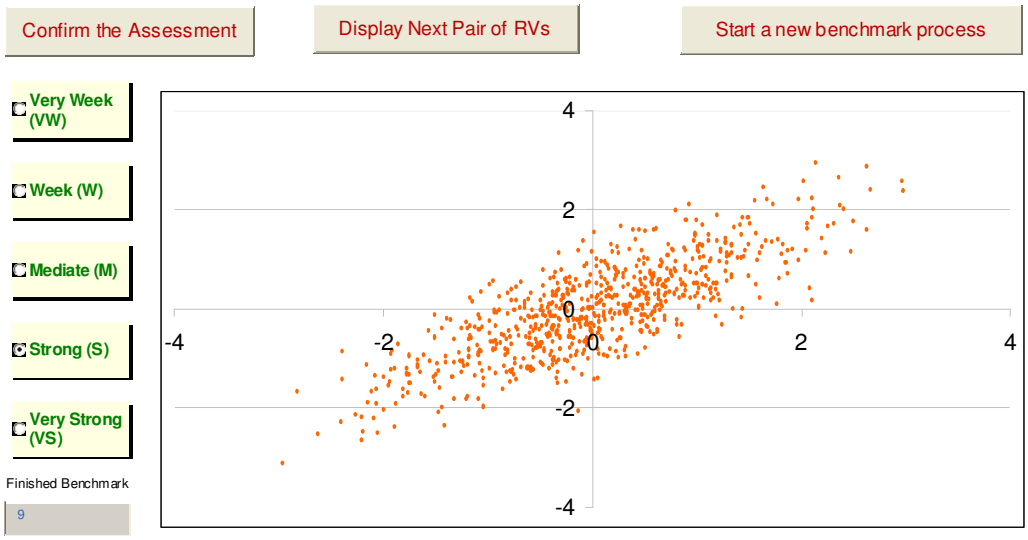


Fig. 3.4 Demonstration of benchmark of the quantitative correlation assessments

By repeating the second step many times with different ρ_{XY}, a table can be built containing the preset correlation coefficient and the expert's qualitative assessment as demonstrated in Table 3.1. The benchmark data are expected to be grouped by the associated qualitative levels as displayed in Fig. 3.5. Statistics such as the mean, the median and the boundaries can then be calculated and used for benchmark of the qualitative assessment of the correlations that are used to build the correlation matrix of a group of input parameters.

Table 3.1 Benchmark table containing the preset correlations and the correlation level assessed by the expert. The qualitative levels are defined as Very weak (VW), Weak (W), Moderate (M), Strong (S), Very strong (VS).

No.	Correlation of the samples on the scatter plot	Qualitative level given by the Expert
1.	0.2	VW
2.	0.8	VS
……		

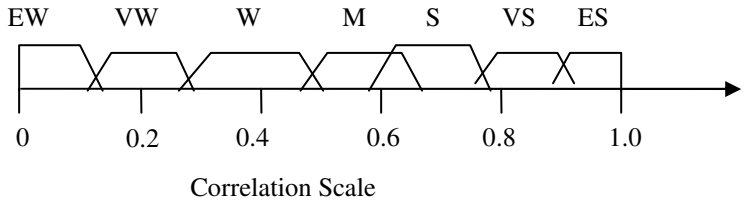


Fig. 3.5 Demonstration of benchmark data grouped by the associated qualitative levels. The two extreme levels Extremely Weak (EW), Extremely Strong (ES) are defined as 0 and 1 for the boundary.

3.11 Define Variance of Occurrence Rate by Eliciting on Waiting Time

Based on the methods developed in the previous sections, assessment of the variance of the input parameter family is required to build the family's covariance matrix. In this section, the method for deriving the variance based on the elicitation is to be developed.

We denote by λ the occurrence rate. We assume that λ follows a truncated normal distribution for $\lambda > 0$, which is defined as

$$\pi(\lambda) = \frac{1}{1-A} Norm(\mu, \sigma_\lambda), \quad \lambda > 0 \tag{3.101}$$

$$A \equiv \Pr(\lambda \le 0) = \int_{-\infty}^{0} Norm(\mu, \sigma_\lambda) d\lambda = \Phi\left(-\frac{\mu}{\sigma_\lambda}\right) \tag{3.102}$$

where μ and σ_λ are the mean and standard deviation of λ respectively; $\Phi(\cdot)$ stands for the standard normal cumulative probability function.

Apparently the term *1-A* in Equation 3.101 works as a normaliser of the probability density function. It is believed that *A* is a small number.

At the stage of uncertainty analysis, the best estimate of the parameters generally has already been done and generally used as the means [Morgan and Henrion 1990]. This is the case with the risk RSSB-SRM with which our case studies will be made. We therefore assume the means are known and focus on the elicitation of the variance.

People's ability to assess the statistical percentiles has been proven in psychological research [Garthwaite, Kadane et al. 2005]. The variance of λ can be derived through eliciting the percentiles of λ. The occurrence rate λ is not observable however. A general principle for elicitation is that we should always elicit an observable variable [Garthwaite, Kadane et al. 2005; O'Hagan, Buck et al. 2006]. Based on this principle, the percentile of the waiting time between two successive occurrences is to be elicited to derive the variance of λ.

Given the occurrence rate λ, the events are assumed to follow a homogeneous Poisson process. Denote the travel miles between the two successive occurrences as *T*. Based on the assumption of homogeneous Poisson process, *T* is exponentially distributed as

$$f(t;\lambda) = \lambda e^{-\lambda t} \tag{3.103}$$

From Equation 3.103 the cumulative probability of *T* given λ can be derived as

$$F(t;\lambda) = \int_0^t f(s;\lambda)ds = 1 - \exp(-\lambda t) \tag{3.104}$$

From Equation 3.101 and 3.104, the predictive cumulative probability of *T* can be derived as

$$\Pr(T < t) = \int_0^{+\infty} F(t;\lambda)\pi(\lambda)d\lambda \tag{3.105}$$

Substituting Equation 3.105 with Equation 3.101 and 3.104 we obtain

$$\begin{aligned}\Pr(T<t) &= \int_0^{+\infty}\left(1-e^{-\lambda t}\right)\frac{Norm(\mu,\sigma_\lambda)}{1-A}d\lambda \\ &= \frac{1}{1-A}\int_0^{+\infty} Norm(\mu,\sigma_\lambda)d\lambda - \frac{1}{1-A}\int_0^{+\infty} e^{-\lambda t}Norm(\mu,\sigma_\lambda)d\lambda \\ &= \frac{1-A}{1-A}-\frac{B}{1-A} \\ &= 1-\frac{B}{1-A}\end{aligned} \tag{3.106}$$

where

$$\begin{aligned}B &= \int_0^{+\infty} e^{-\lambda t}\frac{1}{\sqrt{2\pi}\sigma_\lambda}\exp\left(-\frac{(\lambda-\mu)^2}{2\sigma_\lambda^2}\right)d\lambda \\ &= \int_0^{+\infty}\frac{1}{\sqrt{2\pi}\sigma_\lambda}\exp\left(-\frac{(\lambda-\mu)^2+2\sigma_\lambda^2 t\lambda}{2\sigma_\lambda^2}\right)d\lambda \\ &= \int_0^{+\infty}\frac{1}{\sqrt{2\pi}\sigma_\lambda}\exp\left(-\frac{(\lambda-\lambda_E)^2}{2\sigma_\lambda^2}\right)\cdot\exp\left(\sigma_\lambda^2 t^2/2-\mu t\right)d\lambda \\ &= \exp\left(\sigma_\lambda^2 t^2/2-\mu t\right)\cdot\left[1-\Phi\left(-\frac{\lambda_E}{\sigma_\lambda}\right)\right]\end{aligned} \tag{3.107}$$

where

$$\lambda_E = \mu - t\sigma_\lambda^2 .$$

Define the percentile of the waiting time t_{cp} by

$$cp = \Pr(T \le t_{cp}), \quad cp \in (0,1)$$

Based on Equation 3.106 and 3.107, σ_λ is a monotonic function of t_{cp} for a given $cp \in (0,1)$. The function curve can be solved numerically and used to define σ_λ through eliciting t_{cp}. As generally recommended in the literature, 95 percentile $t_{0.95}$ is to be elicited from the experts [Pearson and Tukey 1965; Keefer and Bodily 1983].

As an example, the precursor family of *road vehicle (RV) incorrectly on LCs and struck by train due to environmental factors* is taken to demonstrate the above method. The family occurrence rate is evaluated as $\mu = 1.18\text{e-}09$ per train travel mile [Dennis 2006]. The occurrence rate per year can be derived through the yearly travel miles, i.e. $M = 275242062\ miles/year$. The curve of $t_{0.95}$ against the candidate σ_λ is plotted in Fig. 3.6. It demonstrates that when σ_λ varies from 0 to $\mu/3.8$, $t_{0.95}$ varies from 111 months to 126 months. The experts are asked to give their belief of $t_{0.95}$ within this interval.

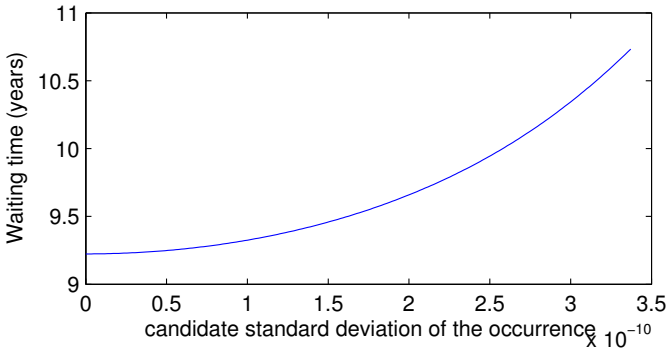


Fig. 3.6 Percentile $t_{0.95}$ versus the candidate standard deviation σ_λ.

3.12 An Example of Building Family Covariance Matrix

In this section we make an example to show the procedure developed above for building the covariance matrix of an input parameter family. In RSSB-SRM version 5, the hazardous event HET-12 is defined as *the derailment of passenger trains* [Dennis 2006]. As summarized in Table 3.2, four precursors related to HET-12, denoted by p_i, $i = 1, \cdots, 4$, are the yearly occurrence rates related to different types of train drivers' errors. Following the procedure defined above, we are going to build the covariance matrix of the four precursors.

Table 3.2 Precursors related to train derailment caused by the train drivers' errors

Precursors Code	Denoted by	Description
POSL----PH	p_1	Overspeeding leading to PT derailment
PSNT----PF	p_2	Severe braking/snatch leading to PT derailment
PSPD----PH	p_3	SPAD at S&C leading to PT derailment
UTRN----UE	p_4	Running into train derailed while in depots/sidings leading to train derailment

Step 1:Define the precursor family and the factors

As summarized in Table 3.2, the four precursors are related to train drivers' errors of different causes. They are also related to the group of the train drivers. By the definition given above, the experts believe that these four precursors form one family and their variance can be derived through the elicitation regarding the four precursors as a group.

One major factor affecting the occurrences of the train drivers' errors is elicited as the drivers monitoring and training program (DMTP). Because all the four precursors are related to the same group of drivers who are given the same monitoring and training program, DMTP is identified as a common factor denoted by X_1. Because there is no difference regarding the DMTP among the four precursors, DMTP is further defined as an invariant factor of the family. The residual uncertainty terms associated with the four precursors are believed to be independent. Therefore, the uncertainty of this family can be defined by the linear model defined in Equation 3.79. We are going to build the family correlation matrix and the covariance matrix according to the procedure developed above.

Step 2:Elicit correlations regarding the factors

We start with eliciting the qualitative correlation assessment between DMTP and the family members, denoted by $r(p_i, X_1),\ i = 1,\cdots,4$. All these correlations are positive. The qualitative assessment are then be transformed to the numerical correlations by the benchmark technique described in Section 3.11. The elicited qualitative correlation assessments and the associated numerical correlations are summarized in Table 3.3.

Table 3.3 Correlations between the family members and the invariant factor driver monitoring and training program

Precursor Codes	Qualitative Correlation Assessment	Correlation number	Denoted by
POSL----PH	VS	0.939	$r(p_1, X_1)$
PSNT----PF	S	0.776	$r(p_2, X_1)$
PSPD----PH	S	0.776	$r(p_3, X_1)$
UTRN----UE	S	0.776	$r(p_4, X_1)$

Step 3:Build the correlation matrix

By the procedure developed above, we will build the family correlation matrices associated with the common factor DMTP and the residual terms respectively. The family correlation matrix can then be assembled by Equation 3.86.

By Equation 3.11, we denote the vector of the precursors in the family by $\mathbf{p} = [p_1, p_2, p_3, p_4]^T$. Based on the correlations given in Table 3.3, the correlation matrix of between the family members and DMT can be defined by Equation 3.54 as

$$r(\mathbf{p}, X_1) = \begin{bmatrix} 0.939 & 0 & 0 & 0 \\ 0 & 0.776 & 0 & 0 \\ 0 & 0 & 0.776 & 0 \\ 0 & 0 & 0 & 0.776 \end{bmatrix}$$

where X_1 stands for the invariant factor DMTP.

We can then build the correlation matrix associated with DMTP by Equation 3.55 as

$$r(\mathbf{p})\big|_{X_1} = r(\mathbf{p}, X_1)\cdot[1]_{4\times4}\cdot r(\mathbf{p}, X_1) = \begin{bmatrix} 0.882 & 0.729 & 0.729 & 0.729 \\ 0.729 & 0.602 & 0.602 & 0.602 \\ 0.729 & 0.602 & 0.602 & 0.602 \\ 0.729 & 0.602 & 0.602 & 0.602 \end{bmatrix}$$

where $[1]_{4\times4}$ is a ones-matrix as defined in Equation 3.25 as

$$[1]_{4\times4} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

We continue to build the family correlation matrix associated with the residual terms. For this purpose, we need to calculate q_i, $i = 1,\cdots,4$, where q_i is the proportions of the i^{th} member's variance that is explained by the uncertainty factors. By Equation 3.71, we have

$$q_1 = r(p_1, X_1)\cdot r(p_1, X_1) = 0.882$$

$$q_2 = r(p_2, X_1)\cdot r(p_2, X_1) = 0.602$$

$$q_2 = r(p_3, X_1)\cdot r(p_3, X_1) = 0.602$$

$$q_2 = r(p_4, X_1)\cdot r(p_4, X_1) = 0.602$$

By Equation 3.75, we have

$$r(\mathbf{p})\big|_{\varepsilon_1} = \begin{bmatrix} 1-q_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0.118 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$r(\mathbf{p})\big|_{\varepsilon_2} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1-q_2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0.398 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$r(\mathbf{p})\big|_{\varepsilon_3} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1-q_3 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0.398 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$r(\mathbf{p})\big|_{\varepsilon_4} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1-q_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.398 \end{bmatrix}$$

By Equation 3.76, we then have

$$r(\mathbf{p})\big|_{\varepsilon} = \sum_{i=1}^{4} r(\mathbf{p})\big|_{\varepsilon_i} = \begin{bmatrix} 0.118 & 0 & 0 & 0 \\ 0 & 0.398 & 0 & 0 \\ 0 & 0 & 0.398 & 0 \\ 0 & 0 & 0 & 0.398 \end{bmatrix}$$

The family correlation matrix can then be assembled by Equation 3.78 as

$$r(\mathbf{p}) = r(\mathbf{p})\big|_{X_1} + r(\mathbf{p})\big|_{\varepsilon} = \begin{bmatrix} 1 & 0.729 & 0.729 & 0.729 \\ 0.729 & 1 & 0.602 & 0.602 \\ 0.729 & 0.602 & 1 & 0.602 \\ 0.729 & 0.602 & 0.602 & 1 \end{bmatrix}$$

Step 4: Derive the standard deviation of the members within the family

We can continue to derive the family covariance matrix through elicitation of the 95 percentile of the waiting time of the family, which is denoted by $t_{0.95}$. The means of the occurrence rates of the precursors in the family are summarized in Table 3.4. The mean of the family occurrence rate is calculated as

$$\mu = 2.34e-9$$

As developed in Section 3.12, we denoted the standard deviation of the family occurrence rate as σ_λ. We calculate $t_{0.95}$ corresponding to two values of σ_λ for the experts' reference when making their assessment. When $\sigma_\lambda = 0$, the family occurrence rate is fixed at the mean and $t_{0.95}$ is calculated by the pure exponential distribution as 55.8 months. When $\sigma_\lambda = \mu/3.8$, $t_{0.95}$ is calculated as 65.3 months. The experts assess $t_{0.95}$ as 59.5 months. The standard deviation of the family occurrence rate can then be derived as 4.69*e*-10. Consequently the standard deviations of the four precursors can be derived by the procedure developed in Section 3.9, which are summarized in Table 3.4.

Table 3.4 Means and the standard deviations of the precursors in the example

Precursors	Mean μ_i	STD σ_i
p_1	3.060E-10	5.480E-11
p_2	1.801E-10	3.906E-11
p_3	1.836E-09	3.982E-10
p_4	1.836E-11	3.982E-12

Step 5: Build the family covariance matrix

With the standard deviations of the precursors derived, by Equation 3.35 we have

$$\boldsymbol{\sigma}_{\mathbf{p}} = \begin{bmatrix} 5.480e-11 & 0 & 0 & 0 \\ 0 & 3.906e-11 & 0 & 0 \\ 0 & 0 & 3.982e-10 & 0 \\ 0 & 0 & 0 & 3.982e-12 \end{bmatrix}$$

Using Equation 3.100, the family covariance matrix can be calculated as

$$\begin{aligned} \operatorname{cov}(\mathbf{p}) &= \boldsymbol{\sigma}_{\mathbf{p}} \cdot r(\mathbf{p}) \cdot \boldsymbol{\sigma}_{\mathbf{p}} \\ &= \begin{bmatrix} 3ee-21 & 1.56e-21 & 1.59e-20 & 1.59e-22 \\ 1.56e-21 & 1.53e-21 & 9.36e-21 & 9.36e-23 \\ 1.59e-20 & 9.36e-21 & 1.59e-19 & 9.54e-22 \\ 1.59e-22 & 9.36e-23 & 9.54e-22 & 1.59e-23 \end{bmatrix} \end{aligned}$$

With the means and the covariance matrix built through elicitation of expert judgement, the joint normal distribution of the precursors is then completely defined.

3.13 Conclusion

In this chapter, a procedure is developed for building the covariance matrix of a group of parameters through elicitation of expert judgement. In a factor analysis, a set of factors are identified as the underlying dependence structure in a correlation matrix. The proposed procedure works as an inverse procedure of the factor analysis. We start with the factors that construct the dependence structure; we then elicit on the factors and derive the correlation matrix.

Through a linear model, the uncertainty in a parameter is structured into the contributions of its uncertainty factors. The correlation between a pair of input parameters is brought by the common factors and the factors in the same class that therefore are correlated. A method has been developed in this chapter to build the correlation by eliciting the correlations regarding the common factors and the

correlated factors. An indicative value is derived as the proportion of the parameter's variance that is explained by the factors. This indicative value cannot exceed 1, which can be used to monitor the assessment during the elicitation. The experts have to compare between the factors and give the consistent assessment that satisfies this constraint. As a result, a positive semi-definite correlation matrix can be guaranteed for multiple input parameters, which is a big feature of this method. Suppose there are *n* parameters having *m* common factors. The number of elicitation parameters is $n \times m$. Compared with directly filling in the correlation matrix of $n(n-1)/2$ cells, this method also features much less elicitation workload.

When a group of parameters have physically similar interpretation, the concept of a parameter family is defined. In risk and reliability analysis most of the input parameters are defined as the occurrence rate of the basic events [Kumamoto and Henley 1996; Modarres 2006]. The sum of the input parameters can therefore be defined as occurrence rate of the family which is meaning to assess the variance. A method is then developed to derive the variance of the input parameters through eliciting the variance of the family occurrence rate. The first benefit of the method is the reduction of the elicitation workload regarding the variance. Secondly the assessment of higher level events is more reliable as is the case with RSSB-SRM.

As the support technique, we select the combination of the qualitative assessment and the benchmark for eliciting the correlation between two random variables. This combination makes the method both easy to implement and rigorous in terms of the probability theory.

At the end of this chapter a method is developed to derive the variance of the occurrence rate by eliciting the percentile of the waiting time between two successive occurrences. The waiting time is an observable variable and people's ability in assessing the statistical percentile has been proved in psychological research and many elicitation practices. Therefore this method is feasible to implement.

From this chapter, we have some work left for future as well. First, we assume a truncated normal distribution for the occurrence rate λ. The assumption is made together with RSSB. We need to investigate if other distributions such as a Gamma distribution are suitable. Furthermore, we need to investigate the possible consequences of adopting a different distribution.

Second, the proposed benchmark method depends on the assumption that experts are able to percept the correlations by observing the scatter plots. For other distributions than normal, the scatter plots may show more complex patterns and mislead experts. Therefore it will be difficult to apply the proposed benchmark method for other types of distributions. This difficult however can be overcome or mitigated by a proper training process. Experts can understand well the patterns associated with a new distribution through the training. With an efficient training procedure, the proposed benchmark method can be used for other distribution. Similarly, the proposed benchmark method can be used for rank correlations as long as an efficient training procedure can be made. We can continue to find some efficient training methods.

Chapter 4

Modelling Uncertainty in Rule Sets of a Safety Event Tree

4.1 Introduction

Event-trees are widely used for modelling the risk scenarios in safety risk analysis. An event tree starts with an initiating event that has the potential to cause safety risk. A set of events follows the initiating event. These events can take different outcomes and affect the final safety risk. Therefore they are called escalation events. A path from the initiating event through all the escalation events is called an accident sequence. The distinct combinations of the outcomes of the escalation events define a finite number of accident sequences of an event tree. A risk scenario is specified at the end of each accident sequence. Associated with each risk scenario, a number of people are supposed to be exposed to the risk of getting injured. The numbers of the people exposed to the risk scenarios are subject to uncertainty. In this research, however, we don't study the uncertainty in these numbers. We focus on studying the uncertainty in the individual injury probabilities associated with the rule sets of the event tree, which is to be discussed next.

The injuries are usually classified into four levels including: no injuries, minor injuries, major injuries and fatalities [Dennis 2006]. Each individual exposed to the risk suffers a probability of getting injured at one of the four levels. We assume that all the individuals exposed to one risk scenario get injured independently. The consequences associated with one risk scenario are defined as the means of injuries at the four levels. Consequently the consequences are calculated as the product of the

number of people on scene and the individual injury probabilities. The escalation events that affect the individual injury probabilities form a subset of the escalation events. A rule set is defined when each escalation event in the subset takes a specific outcome. Multiple risk scenarios can be associated with the same rule set. Once the individual injury probabilities of the rule set are defined, they can be used for all the associated risk scenarios. Therefore the workload for populating the risk model can be reduced by modelling on the rule sets.

There are still difficulties in modelling the uncertainty in the rule sets of an event tree composing many escalation events. First, the rule sets are intensively interwoven to each other through the escalation events. It makes it very difficult to directly assess the correlations among rule sets. Second, the number of the rule sets increases exponentially with the number of the hazards, which implies too many pairwise correlations to assess for a large Event-tree model. Third, the individual injury probabilities should have a monotonicity property: a rule set associated with worse outcomes of all the escalation events should have individual injury probabilities not less than another rule set associated with better outcomes. For instance, two rule sets in [Dennis 2006] are defined for the collision of a passenger train and a road vehicle with high and low approach speeds respectively. By considering the physical forces involved, when all the other situations are the same, the individual injury probabilities due to the collision associated with high approach speed should not be less than those associated with low approach speed. Such an order relationship cannot be modelled purely by the pairwise correlations among the rule sets.

Aimed at the above difficulties, we identify along the sequence of the escalation events the hazard types that are the materials or activities with potential to cause injuries. Depending on the outcomes of the escalation events associated with a risk scenario, one hazard type is defined at a specific level, which is called a hazard. As a result, a rule set is defined by a sequence of hazards decided by the outcomes of the escalation events associated with each risk scenario. These concepts are demonstrated with an example regarding the risk of the possible rupture of a tank containing compressed natural gas (CNG). As summarized in Table 4.1, five distinct

rule sets are defined by the combinations of the hazards that are defined on the outcomes of the escalation events. This example is to be discussed in details later on in this chapter.

Table 4.1 Example of the rule sets decided by the outcomes of the escalation events

	Outcomes of the escalation events			**Hazards**	
No.	**Release Mode**	**Ignition Mode**	**Dispersion Type**	**Explosion**	**Fire**
1.	Instantaneous	Immediate	N/A	Explosion I	Fireball
2.	Instantaneous	Delayed	Dense cloud	Explosion II	Flash fire I
3.	Instantaneous	Delayed	Buoyant	Explosion III	Flash fire II
4.	Instantaneous	No	N/A	No	No
5.	Gradual	Immediate	N/A	No	Jet flame
6.	Gradual	Delayed	Dense cloud	Explosion II	Flash fire I
7.	Gradual	Delayed	Buoyant	Explosion III	Flash fire II
8.	Gradual	No	N/A	No	No

We continue to define the injury probabilities of a rule set through the definitions of the hazards. We assume that, given the sequence, the hazards are independent in terms of causing the injuries to the people exposed to the risk scenario. When an individual person caught in the risk scenario is not lucky enough to escape all the hazards, the person becomes part of the consequence. After each hazard in the sequence, an individual exposed to the risk scenario either keeps the same level of injury or suffers a higher level injury. We assume that the transition probabilities associated with a hazard depend only on the person's injury level before the hazard. Consequently, the individual injury probabilities of a rule set can be defined on the associated hazards by a Markov Chain model [Ross 2003]. To build up the Markov Chain model, we need to elicit the transition probabilities associated with each hazard. Once the transition probabilities are elicited for one hazard, they can be used

for all the rule sets containing the same hazard. Therefore, modelling on the hazards can significantly reduce the elicitation work.

For modelling the uncertainty in the rule sets, we can then model the uncertainty in the transition probabilities. We assume that the transition probabilities associated with different hazard types to be independent. It means that a given transition probability of one hazard does not tell any information on the transition probability of the hazards of another type. We assume that the transition probabilities associated with the same hazard type are proportional. By the assumption, an injury atom is defined for each hazard type associated with the worst hazard level. All the transition probabilities associated with the same hazard type are then defined on the injury atom. The individual injury probabilities of the rule sets can then modelled on the injury atoms. The dependences among the rule sets are modelled automatically through the injury atoms. The order relationship can also be kept between two rule sets associated with general worse and better hazards respectively.

In most of the cases, we are more concerned with the major injuries and the fatalities than with the minor injuries. We need also to keep the elicitation work and time reasonable to the experts. For these two purposes, we develop a model of two levels of injuries as a simplified case of the model of four levels of injuries. A demonstration example of the model of two levels of injuries is then made on three rule sets of RSSB-SRM [Dennis 2006].

4.2 Risk Scenarios and Consequences

An Event-tree starts with an initiating event that has the potential to cause the safety risk. Shown in Fig. 4.1 is an event-tree example that is extracted from [Modarres 2006]. The initiating event in the example is the rupture of the tank containing compressed natural gas (CNG). After the initiating event, there can be many other events that can have different outcomes and affect the final consequence. These following events are accordingly called escalation events [Marsh and Bearfield 2008]. In the event tree shown in Fig. 4.1, five escalation events are included as (1) gas

release mode; (2) expansion and ignition; (3) dispersion type: (4) fire type; and (5) fire location. Each escalation event is represented by a node; the outcomes of an escalation event are represented by the branches following the node. Given the outcomes of the previous events, the outcomes of one event may not affect the final consequence. In this case, the event is set as not applicable (N/A) that is a special outcome; and there is one branch following the node. Such cases can be found in the example shown in Fig. 4.1.

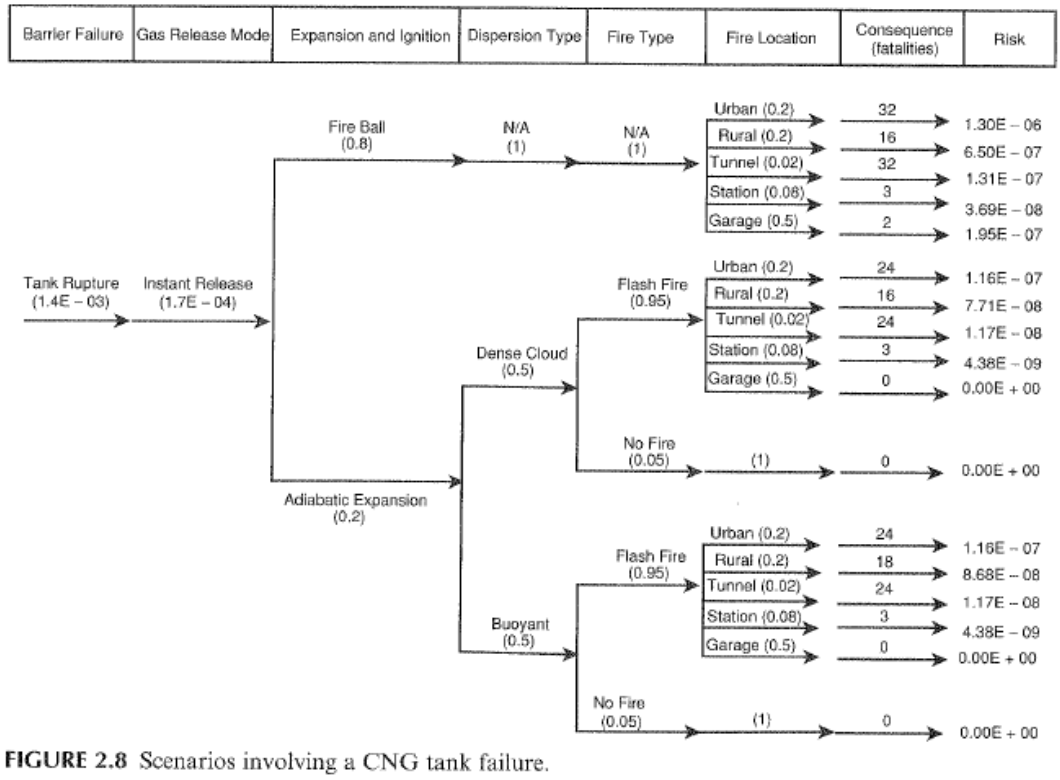


FIGURE 2.8 Scenarios involving a CNG tank failure.

Fig. 4.1 An Event-tree example extracted from [Modarres 2006].

We denote the escalation events in the same sequence as in the event tree by $E_1, \cdots, E_i, \cdots, E_L$, where L represents the number of the escalation events. We denote the outcomes of the event E_i, typically yes or no, as

$$\mathscr{W}(E_i) = \{\omega_{i,j}\},\ j = 1, \cdots, K_i$$

where K_i stands for the number of the outcomes of E_i.

Following [Papazoglou 1998], the set $\mathscr{W}(E_i)$ is called the outcome space of the event E_i. All the possible combinations of the outcomes $\omega_{i,j}$, $i = 1, \cdots, L$, define the outcome space of the escalation events $E_1, \cdots, E_i, \cdots, E_L$, which is denoted by

$$\mathscr{W}(E_1, \cdots, E_i, \cdots, E_L) = \mathscr{W}(E_1) \otimes \cdots \otimes \mathscr{W}(E_i) \otimes \cdots \otimes \mathscr{W}(E_L) \tag{4.1}$$

where $\otimes$ stands for the Cartesian multiplication.

With each escalation event having a specific outcome, a path from the initiating event through all the escalation events resulting in the final consequence is called an accident sequence. A finite number of accident sequences of an event tree are defined by the distinct combinations of the outcomes of the escalation events. For instance, there are 18 scenarios defined in the example shown in Fig. 4.1. A specific risk scenario is defined at the end of each accident sequence. We denote the risk scenario associated with the ℓ^{th} accident sequence as S_ℓ. We denote the outcome of E_i associated with S_ℓ as $e(\ell, i)$

$$e(\ell, i) \in \mathscr{W}(E_i) \tag{4.2}$$

Consequently the risk scenario S_ℓ can then be defined as

$$S_\ell = (e(\ell,1), \cdots, e(\ell,i), \cdots, e(\ell,L)) \tag{4.3}$$

By Equation 4.1, we have

$$(e(\ell,1), \cdots, e(\ell,i), \cdots, e(\ell,L)) \in \mathscr{W}(E_1, \cdots, E_i, \cdots, E_L)$$

Therefore each risk scenario corresponds to an element of the outcome space of the escalation events.

With the outcomes of all the escalation events defined, the consequence due to a risk scenario can then be defined. Usually the injuries are classified into four levels including: no injury, minor injury, major injury, and fatality. When an individual is exposed to a risk scenario, the probabilities of the individual getting injured at the four levels are called the individual injury probabilities. We denote the individual injury probabilities associated with S_ℓ as $\pi_\ell(m)$, where $m=1,\cdots,4$ stand for no injury, minor injury, major injury, and fatality respectively. Therefore we have

$$\sum_{m=1}^{4}\pi_\ell(m)=1,\ \pi_\ell(m)\geq 0 \tag{4.4}$$

In addition to the individual injury probabilities, a number of people are supposed to be exposed to a risk scenario. The number of people exposed to the risk scenario is affected by the outcomes of associated escalation events. For instance, the event *Fire Location* as shown in Fig. 4.1 has five outcomes including: Urban (0.2), Rural (0.2), Tunnel (0.02), Station (0.08), and Garage (0.5). The numbers in the parenthesis are the probabilities of the associated outcomes. The mean of the number of the people exposed to the scenario is set for each location. Another example can be found in RSSB-SRM [Dennis 2006]. According to the time intervals during a day, the loading type of the passenger trains in the RSSB-SRM has four outcomes including Night loading (0.1), Off-peak loading (0.72), Peak loading (0.135), and Crush loading (0.045). The mean of the number of the passengers onboard is set for each loading type. In summary, the mean of the number of people exposed to the risk scenario is decided by the outcomes of the associated escalation events. We denote the mean of the number of people exposed to the risk scenario S_ℓ as N_ℓ.

The consequences associated with the risk scenario are defined as the numbers of the injuries at the four levels. We assume that all the individuals exposed to one risk

scenario get injured independently. We denote the consequences associated with S_ℓ as $c_\ell(m)$, where $m = 1,\cdots,4$ stand for no injury, minor injury, major injury, and fatality respectively. The consequences of a risk scenario are defined as the means of injuries at the four levels as

$$c_\ell(m) = N_\ell(P) \times \pi_\ell(m),\ m = 1,\cdots,4 \tag{4.5}$$

By applying Equation 4.4 with application 4.5, we have

$$\sum_{m=1}^{4} c_\ell(m) = N_\ell \tag{4.6}$$

By Equation 4.5, the uncertainty in $c_\ell(m)$ is driven by: (1) the uncertainty in N_ℓ, i.e. the number of the people exposed to S_ℓ; and (2) the uncertainty in $\pi_\ell(m)$, i.e. the individual injury probability. In this research, we concentrate on developing the methods for modelling the uncertainty in $\pi_\ell(m)$ and treat N_ℓ as a constant.

4.3 Rule Sets

A hazard type is either a material or activity that has the potential to cause injuries to the people exposed to the risk scenario. Some common hazard types in industry sectors include explosion, fire, and toxic releases etc as summarized in [Andrews and Moss 2002]. On the railway system three common hazard types are identified as the mechanical impact, the fire, and the toxic release [Dennis 2006]. Along the escalation event sequence $(E_1,\cdots,E_i,\cdots,E_L)$, we can identify a sequence of hazard types, which is defined as

$$\mathbf{T} = (T_1,\cdots,T_j,\cdots,T_D) \tag{4.7}$$

where T_j represents the j^{th} hazard type; D stands for the number of the hazard types.

Each hazard type is affected by a subset of escalation events. We define the subset of escalation events affecting T_j as

$$\mathscr{E}(T_j) = \{E_i | E_i \text{ affects the hazard level of } T_j\} \tag{4.8}$$

For example, the first hazard type included in RSSB-SRM [Dennis 2006] is the mechanical impact in the collision of a passenger train with a road vehicle. By considering the physical forces involved, the approach speed of the train affects the individual casualty probability due to the mechanical impact. One escalation event included in RSSB-SRM [Dennis 2006] is the type of the level crossing. Different limits of the approach speed are imposed on the train on different types of level crossings [RSSB 2004]. Therefore the level crossing type is one escalation event that affects the mechanical impact hazard type. Another hazard type included in RSSB-SRM [Dennis 2006] is fire. It is affected by the escalation event defining whether there are extra flammable materials involved in the accident. The third hazard type in RSSB-SRM is the toxic goods, which is affected by the escalation events defining the amount of the toxic released in the accident.

As defined in Equation 4.3, a risk scenario S_ℓ is defined as a sequence of outcomes of the escalation events along the event tree. The hazard types defined in Equation 4.7 are accordingly defined at the specific levels. We define a specific level of a hazard type as a hazard. We denote by $H_{\ell,j}$ the hazard of the type T_j specified by the risk scenario S_ℓ. Corresponding to $\mathscr{E}(T_j)$ defined in Equation 4.8, we define the set of the outcomes associated with $H_{\ell,j}$ as

$$\mathscr{O}(H_{\ell,j}) = \{e(\ell,i) | E_i \in \mathscr{E}(H_j)\} \tag{4.9}$$

where $e(\ell,i) \in \mathscr{W}(E_i)$ stands for the outcome of E_i associated with S_ℓ.

Associated with S_ℓ, a sequence of *D* hazards corresponding to the *D* hazard types in $\mathbf{T}$ is defined as

$$R_\ell = \left(H_{\ell,1},\cdots,H_{\ell,j},\cdots,H_{\ell,D}\right) \tag{4.10}$$

We assume that the sequence of the hazards define the individual injury probabilities of an individual exposed to the risk. Consequently R_ℓ is called a rule set [Dennis 2006].

Based on the definition in Equation 4.10, when $H_{\ell(1),j} = H_{\ell(2),j}$, $j = 1,\cdots,D$, the rule sets $R_{\ell(1)}$ and $R_{\ell(2)}$ are identical, which is denoted by

$$R_{\ell(1)} = R_{\ell(2)} \tag{4.11}$$

Therefore multiple risk scenarios can be associated with the same rule sets.

We have assumed that the sequence of the hazards define the individual injury probabilities of a rule set. Therefore, the identical rule sets $R_{\ell(1)} = R_{\ell(2)}$ have the same individual injury probabilities, which are denoted as

$$\pi_{\ell(1)}(m) = \pi_{\ell(2)}(m),\ m = 1,\cdots,4 \tag{4.12}$$

Once the individual injury probabilities of the rule set are assessed, they can be used for all the other risk scenarios associated with the same rule set. It represents an important benefit of defining the consequences through the rule sets by Equation 4.5.

We continue to demonstrate the above definitions regarding the rule sets with an example. We use the system on the rupture of the tank containing compressed natural gas as shown in Fig. 4.1. In this system, two hazard types are identified as: H_1: explosion; and H_2: fire [Modarres 2006]. The levels of H_1 and H_2 are decided by

the outcomes of three escalation events including: (1) E_1: gas release mode; (2) E_2: ignition model; (3) E_3: dispersion type. The escalation event E_1 has two outcomes defined as: (1) $\omega_{1,1}$: instantaneous release; (2) $\omega_{1,2}$: gradual release. The escalation event E_2 has three outcomes defined as: (1) $\omega_{2,1}$: immediate ignition; (2) $\omega_{2,2}$: delayed ignition; (2) $\omega_{2,3}$: no ignition. The escalation event E_3 has two outcomes defined as: (1) $\omega_{3,1}$: dense cloud; (2) $\omega_{3,2}$: buoyant. Associated with one risk scenario, the three escalation events take the specific outcomes and consequently H_1 and H_2 take the specific hazard. The corresponding hazards and the associated outcomes of the escalation events are summarized in Table 4.1. Associated with H_1 there are three hazards defined as: (1) Explosion at Level I; (2) Explosion at Level II; and (3) Explosion at Level III. Associated with H_2, there are four hazards defined as: (1) Fireball; (2) Flash fire at Level I; (3) Flash fire at Level II; and (4) Jet blame.

As summarized in Table 4.1, we can see that there are five distinct rule sets defined by the combinations of the hazards as

R_1: explosion at Level I, fireball

R_2: explosion at Level II, flash fire at Level I

R_3: explosion at Level III, flash fire at Level II

R_4: no explosion, Jet flame

R_5: no explosion, no fire

This example shows that all the risk scenarios can be associated with the five rule sets, and hence that the rule set concept offers potential for a significant reduction of the elicitation work.

In this section, we have defined the rule sets for the risk scenarios. We have shown that multiple risk scenarios can be associated with the same rule set. Once the individual injury probabilities of a rule set are assessed, they can be used for all the associated risk scenarios. Therefore defining the consequences through the rule sets

by Equation 4.5 requires less elicitation work than assessing the risk scenarios directly. However for a large event tree of many hazard types, there are still too many rule sets to assess through elicitation. As discussed above, the rule sets are defined on a sequence of hazards decided by the outcomes of the escalation events. Consequently there are three difficulties in modelling the uncertainty in the rule sets as described above as: (1) the difficulty in directly assessing the correlations among intensively interwoven rule sets; (2) too many pairwise correlations to assess for a large Event-tree model; (3) the difficulty in keeping the monotonicity property in the individual injury probabilities. Due to the construction process of the event trees, the above difficulties are generic when we need to model the uncertainty in the rule sets. To solve the difficulties, we are going to define the individual injury probabilities of a rule set on the associated hazards by a Markov Chain model [Ross 2003]. Therefore once a hazard is assessed, it can be used in all the other rule set containing the hazard.

4.4 Definition of Rule Sets on Hazards

As described before, a rule set is defined by a sequence of hazards decided by the outcomes of the escalation events. We assume that, given the sequence, the hazards are independent in terms of causing the injuries to the people exposed to the risk scenario. When an individual caught in the risk scenario is not lucky enough to escape all the hazards, the person becomes part of the consequence. After each hazard in the sequence, an individual exposed to the risk scenario is either kept at the existing level of injury or moved to a higher level injury, as shown in Fig. 4.2. We assume that the probabilities of a hazard moving a person to higher levels of injuries depend only on the person's existing injury level before the hazard. Consequently, the individual injury probabilities of a rule set can be defined on the associated hazards by a Markov Chain model [Ross 2003].

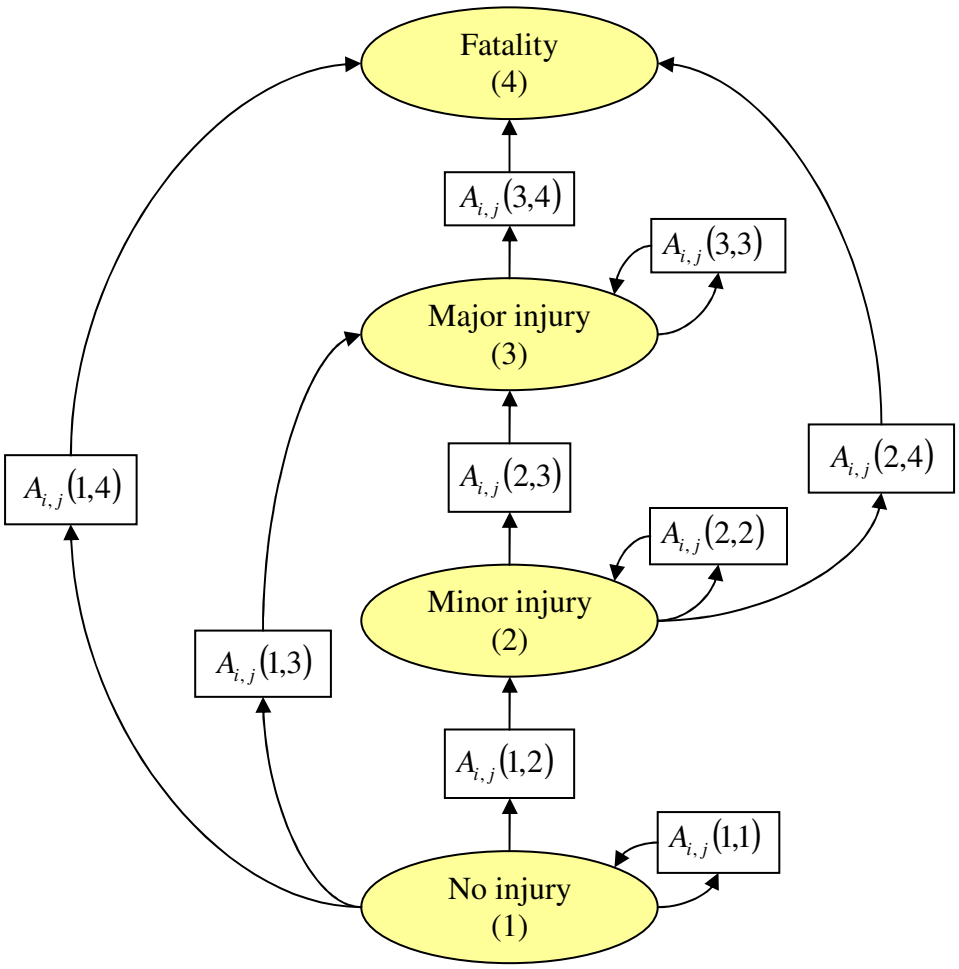


Fig. 4.2 Diagram of the Markov Chain model of the individual injuries

As shown in Fig. 4.2, we denote by $A_{i,j}(m,k)$ the hazard $H_{i,j}$'s transition probability from m level to k level, where $1 \le m \le k \le 4$ stand for the four levels of injuries as defined before. Clearly we have

$$0 \le A_{i,j}(m,k) \le 1,\ 1 \le m \le k \le 4 \tag{4.13}$$

We denote the individual injury probabilities of the rule set R_i before and after $H_{i,j}$ as $\pi_{i,j-1}(m)$ and $\pi_{i,j}(m)$ respectively, where $m = 1,\cdots,4$ stand for the four levels of

injuries respectively. The individual injury probabilities $\pi_{i,j-1}(m)$ and $\pi_{i,j}(m)$ are connected by the transition probabilities $A_{i,j}(m,k)$ as shown in Fig. 4.3.

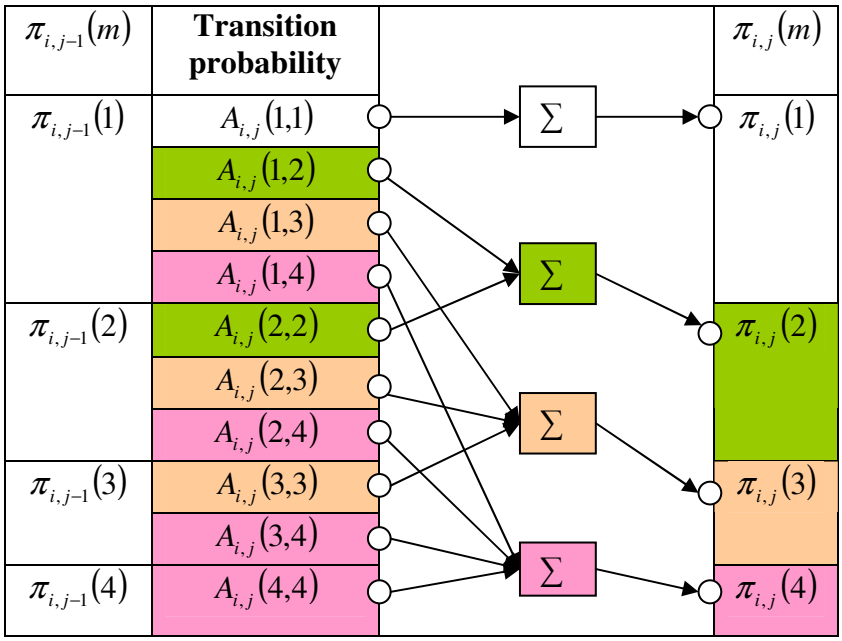


Fig. 4.3 Individual injury probabilities connected by the transition probabilities

As shown in Fig. 4.3, to have no injury after the hazard $H_{i,j}$, the individual must have no injury before $H_{i,j}$ and follow the transition $A_{i,j}(1,1)$. Therefore we have

$$\pi_{i,j}(1) = \pi_{i,j-1}(1) \cdot A_{i,j}(1,1) \tag{4.14}$$

In the same way, we have

$$\pi_{i,j}(2) = \pi_{i,j-1}(1) \cdot A_{i,j}(1,2) + \pi_{i,j-1}(2) \cdot A_{i,j}(2,2) \tag{4.15}$$

$$\pi_{i,j}(3) = \pi_{i,j-1}(1) \cdot A_{i,j}(1,3) + \pi_{i,j-1}(2) \cdot A_{i,j}(2,3) + \pi_{i,j-1}(3) \cdot A_{i,j}(3,3) \tag{4.16}$$

$$\pi_{i,j}(4)=\pi_{i,j-1}(1)\cdot A_{i,j}(1,4)+\pi_{i,j-1}(2)\cdot A_{i,j}(2,4)+\pi_{i,j-1}(3)\cdot A_{i,j}(3,4)+\pi_{i,j-1}(4)$$

(4.17)

The above four equations can be summarized in one as

$$\pi_{i,j}(m)=\sum_{k=1}^{m}\pi_{i,j-1}(k)\cdot A_{i,j}(k,m),\ m=1,\cdots,4 \tag{4.18}$$

We call $A_{i,j}(m,m)$ the dummy transition probability at the *m* level of injury. We will elicit the transition probabilities $A_{i,j}(m,k)$, $m<k$, and determine the dummy probabilities using

$$A_{i,j}(m,m)=1-\sum_{k=m+1}^{4}A_{i,j}(m,k),\ m=1,\cdots,4 \tag{4.19}$$

Applying Equation 4.19 with Equation 4.18, we have

$$\pi_{i,j}(m)=\sum_{k=1}^{m-1}\pi_{i,j-1}(k)\cdot A_{i,j}(k,m)+\pi_{i,j-1}(m)\cdot\left(1-\sum_{j=m+1}^{4}A_{i,j}(m,j)\right) \tag{4.20}$$

We assume that the people exposed to the risk scenario are initially uninjured. Therefore we have

$$\pi_{i,0}(1)=1,\quad \pi_{i,0}(2)=\pi_{i,0}(3)=\pi_{i,0}(4)=0 \tag{4.21}$$

By Equations 4.20 and 4.21, the individual injury probabilities of a rule set can be defined when the transition probabilities are assessed for all the hazards associated with the rule set. Suppose that the rule set has *D* hazards in sequence, the individual injury probabilities of the rule set R_i are then defined by

$$\pi_i(m) = \pi_{i,D}(m),\ m = 1,\cdots,4 \tag{4.22}$$

We have so far defined the individual injury probabilities of the rule set through a Markov Chain model. Once the transition probabilities associated with the hazard $H_{i(1),j}$ are assessed, they can be used for any other rule set $R_{i(2)}$ when $H_{i(1),j} = H_{i(2),j}$. It means a significant reduction of the elicitation work in populating the model. We will continue to model the uncertainty in the individual injury probabilities of the rule sets.

4.5 Uncertainty in Individual Injury Probabilities of Rule Sets

4.5.1 Uncertainty in Transition Probabilities

We have so far define the individual injury probabilities of the rule set on the transition probabilities of the associated hazards. The transition probabilities however cannot be assessed with certainty. Therefore the transition probability $A_{i,j}(m,k)$ is defined as a random variable. We denote the mean of $A_{i,j}(m,k)$ by

$$\varphi_{i,j}(m,k) = E\left[A_{i,j}(m,k)\right],\ 1 \le m < k \le 4 \tag{4.23}$$

Clearly, we have

$$0 \le \varphi_{i,j}(m,k) \le 1 \tag{4.24}$$

The uncertainty in $A_{i,j}(m,k)$ belongs to epistemic uncertainty category. It needs to be elicited from experts. After that, the uncertainty in individual injury probabilities of the rule set can be derived.

To model the uncertainty in the transition probabilities, we need to discuss the relations among them. A given transition probability associated with one hazard type does not tell the expert any information about the transition probabilities associated

with other hazard types. Therefore we assume that the transition probabilities associated with different hazard types are independent.

Toward the same hazard type, however, experts are either too optimistic or too pessimistic. If an expert gives an assessment larger than the mean of one transition probability, he will do the same to another transition probability of the same hazard type. This has been agreed with the experts in RSSB. The Dirichlet distribution is not able to keep such a property; consequently it is not suitable for modelling the uncertainty in the transition probabilities associated with the same hazard type.

For simplification, we assume that the expert's assessments of the transition probabilities of the same hazard type deviate from the associated means proportionally. Based on the assumption, the transition probabilities associated with the hazard type T_j have the form:

$$\frac{A_{i,j}(m,k)}{\varphi_{i,j}(m,k)} = \frac{A_j}{\gamma_j}, \text{ for all } i = 1,\cdots,n,\ 1 \le m < k \le 4 \tag{4.25}$$

where n stands for the number of the rule sets; A_j is a random variable; $\gamma_j = E(A_j)$ is the mean of A_j.

In Equation 4.25, we set $1 \le m < k \le 4$ to exclude dummy transition probabilities from the assumption of proportional transition probabilities. After the uncertainty in the real transition probabilities $A_{i,j}(m,k)$, $1 \le m < k \le 4$, are elicited, the uncertainty in the dummy transition probabilities is automatically defined by Equation 4.19.

Based on Equation 4.25, the uncertainties of two transition probabilities associated with the same hazard type can be modelled with copulas [Bedford and Cooke 2001]. As a more convenient way for multiple transition probabilities associated with the same hazard type, however, we are going to define the injury atom for each hazard type.

Equation 4.25 shows that all the transition probabilities associated with T_j can be defined when A_j, γ_j and $\varphi_{i,j}(m,k)$'s are elicited. Therefore A_j works as a reference and is called the injury atom associated with T_j. The injury atom A_j however does not have a meaning definition so that it can be elicited. It needs to be associated with some meaningful parameter for elicitation. By Equation 4.25, any transition probability $A_{i,j}(m,k)$ seems a choice for the injury atom A_j. Unfortunately it is not because the injury atom A_j must be defined to satisfy two constraints. First, by the inequality 4.13, we have

$$0 \le A_{i,j}(m,k) \le 1,\ 1 \le m < k \le 4$$

Second, we define

$$A_{i,j}(m) = \sum_{h=m+1}^{4} A_{i,j}(m,h),\ 1 \le m < 4 \tag{4.26}$$

We can see that $A_{i,j}(m)$ measures the transition probability from the *m*-level injury to all the higher levels of injuries. We call $A_{i,j}(m)$ the general transition probabilities from the *m*-level injury. Clearly we have

$$0 \le A_{i,j}(m) \le 1,\ i = 1,\cdots,n,\ 1 \le m < 4 \tag{4.27}$$

The injury atom A_j must be defined to satisfy these two constraints.

4.5.2 Injury Atom

Based on Equation 4.25, the injury atom A_j needs to be defined so that the uncertainty in the transition probabilities of the same hazard type T_j can be defined.

We need to associated A_j with some meaningful parameters so that it can be elicited. Furthermore, the definition of A_j must satisfy two constraints as imposed by the inequalities 4.13 and 4.27.

We start with discussing the transition probabilities associated with one rule set R_i. From Equation 4.25, the transition probabilities from the *m*-level injury satisfy

$$\frac{A_{i,j}(m,m+1)}{\varphi_{i,j}(m,m+1)} = \cdots = \frac{A_{i,j}(m,k)}{\varphi_{i,j}(m,k)} = \cdots = \frac{A_{i,j}(m,4)}{\varphi_{i,j}(m,4)} = \frac{A_j}{\gamma_j} \tag{4.28}$$

From Equation 4.28, we have

$$\frac{A_{i,j}(m,k)}{\varphi_{i,j}(m,k)} = \frac{\sum_{h=m+1}^{4} A_{i,j}(m,h)}{\sum_{h=m+1}^{4} \varphi_{i,j}(m,h)} = \frac{A_j}{\gamma_j},\ k = m+1,\cdots,4 \tag{4.29}$$

Corresponding to the definition in Equation 4.26, we define

$$\gamma_{i,j}(m) = \sum_{h=m+1}^{4} \varphi_{i,j}(m,h) \tag{4.30}$$

By Equations 4.26, 4.23 and 4.30, we have

$$E\left(A_{i,j}(m)\right) = \sum_{h=m+1}^{4} E\left(A_{i,j}(m,h)\right) = \gamma_{i,j}(m) \tag{4.31}$$

Therefore $\gamma_{i,j}(m)$ is the mean of $A_{i,j}(m)$.

By applying Equations 4.26 and 4.30 with Equation 4.29, we have

$$\frac{A_{i,j}(m,k)}{\varphi_{i,j}(m,k)} = \frac{A_{i,j}(m)}{\gamma_{i,j}(m)} = \frac{A_j}{\gamma_j}, \tag{4.32}$$

By Equation 4.32, the injury atom A_j can be associated with the general transition probability $A_{i,j}(m)$ that can be elicited and can keep the inequalities 4.13 and 4.27. The injury atom A_j, however, needs to be defined to for all the rule sets. We will discuss the situation of multiple rule sets next.

Suppose that $R_{i(1)}$ and $R_{i(2)}$ are two rule sets, where $i(1), i(2) = 1, \cdots, n$. Associated with the hazard type T_j, the two hazards defined for $R_{i(1)}$ and $R_{i(2)}$ are denoted as $H_{i(1),j}$ and $H_{i(2),j}$ respectively. Suppose $A_{i(1),j}(m_1, k_1)$ and $A_{i(2),j}(m_2, k_2)$ are two transition probabilities associated with $H_{i(1),j}$ and $H_{i(2),j}$ respectively, where $1 \le m_1 < k_1 \le 4$, $1 \le m_2 < k_2 \le 4$. Using Equations 4.32, we have

$$\frac{A_{i(1),j}(m_1,k_1)}{\varphi_{i(1),j}(m_1,k_1)} = \frac{A_{i(1),j}(m_1)}{\gamma_{i(1),j}(m_1)} = \frac{A_j}{\gamma_j} \tag{4.33}$$

and

$$\frac{A_{i(2),j}(m_2,k_2)}{\varphi_{i(2),j}(m_2,k_2)} = \frac{A_{i(2),j}(m_2)}{\gamma_{i(2),j}(m_2)} = \frac{A_j}{\gamma_j} \tag{4.34}$$

Therefore, we have

$$\frac{A_{i(1),j}(m_1,k_1)}{\varphi_{i(1),j}(m_1,k_1)} = \frac{A_{i(1),j}(m_1)}{\gamma_{i(1),j}(m_1)} = \frac{A_{i(2),j}(m_2)}{\gamma_{i(2),j}(m_2)} = \frac{A_{i(2),j}(m_2,k_2)}{\varphi_{i(2),j}(m_2,k_2)} = \frac{A_j}{\gamma_j} \tag{4.35}$$

It shows that the general transition probabilities are proportional as well when all the transition probabilities associated with the same hazard type are assumed to be

proportional. From Equation 4.35, we continue to develop the definition of the injury atom A_j for the hazard type T_j.

From Equation 4.35, we then have

$$A_{i(1),j}(m_1) = \frac{\gamma_{i(1),j}(m_1)}{\gamma_{i(2),j}(m_2)} A_{i(2),j}(m_2) \tag{4.36}$$

Suppose

$$0 \le \gamma_{i(1),j}(m_1) \le \gamma_{i(2),j}(m_2) \le 1 \tag{4.37}$$

Consequently we have

$$\frac{\gamma_{i(1),j}(m_1)}{\gamma_{i(2),j}(m_2)} \le 1 \tag{4.38}$$

Applying the inequality 4.38 with Equation 4.36, we have

$$A_{i(1),j}(m_1) \le A_{i(2),j}(m_2) \tag{4.39}$$

Recalling that $0 \le A_{i,j}(m,k) \le 1$. By Equation 4.26, we have

$$A_{i(1),j}(m_1, k_1) \le A_{i(1),j}(m_1) \tag{4.40}$$

Applying the inequality 4.40 with the inequality 4.39, we have

$$A_{i(1),j}(m_1, k_1) \le A_{i(1),j}(m_1) \le A_{i(2),j}(m_2) \tag{4.41}$$

Because $A_{i(1),j}(m_1,k_1)$ and $A_{i(2),j}(m_2,k_2)$ can be any two transition probabilities associated with the hazard type T_j, we can safely generalize the above equations and inequalities to all the transition probabilities associated with T_j. According to the inequality 4.37, we define

$$\gamma_{i(M),j}(m_M) = \max_{\substack{i=1,\cdots,n \\ m=1,\cdots,3}} \gamma_{i,j}(m) \tag{4.42}$$

where *n* stands for the number of the rule sets of the event tree; $i(M)$ and m_M are the indices of the rule set and the injury level associated with the maximum general transition probability respectively.

According to the inequality 4.41, we have

$$A_{i,j}(m,k) \le A_{i,j}(m) \le A_{i(M),j}(m_M), \text{ for all } i=1,\cdots,n,\ 1 \le m < k \le 4 \tag{4.43}$$

Based on the inequality 4.43, the inequalities 4.13 and 4.27 can be satisfied when A_j is associated with $A_{i(M),j}(m_M)$. Therefore, we define

$$A_j = A_{i(M),j}(m_M) \tag{4.44}$$

and

$$\gamma_j = \gamma_{i(M),j}(m_M) \tag{4.45}$$

By Equations 4.44 and 4.45, we have

$$E(A_j) = \gamma_j \tag{4.46}$$

By Equations 4.42 and 4.44, the injury atom A_j can be associated with $A_{i(M),j}(m_M)$, which is called the worst general transition probability. Therefore, the uncertainty in the transition probabilities associated with T_j can be built in three steps. First, we elicit the means of all the transition probabilities, i.e. $\varphi_{i,j}(m,k)$, $i=1,\cdots,n$, $1 \le m < k \le 4$. Second, we search for the worst general transition probability, i.e. the one of the maximum mean $\gamma_j = \gamma_{i(M),j}(m_M)$. Third, we elicit the distribution of $A_j = A_{i(M),j}(m_M)$, for which the method will be discussed later on. All the transition probabilities associated with T_j can then be defined by Equation 4.32. In this way, the inequality 4.13 and 4.27 can be satisfied.

4.5.3 Uncertainty in Individual Injury Probabilities through Injury Atoms

With A_j and γ_j elicited, using Equation 4.32, we have

$$A_{i,j}(m,k) = \frac{\varphi_{i,j}(m,k)}{\gamma_j} A_j,\ i=1,\cdots,n,\ 1 \le m < k \le 4 \tag{4.47}$$

According to Equation 4.47, we define

$$z_{i,j}(m,k) = \frac{\varphi_{i,j}(m,k)}{\gamma_j} \tag{4.48}$$

From Equations 4.47 and 4.48, we have

$$A_{i,j}(m,k) = z_{i,j}(m,k) \cdot A_j \tag{4.49}$$

We call $z_{i,j}(m,k)$ the weight of the transition probability $A_{i,j}(m,k)$ relative to the injury atom A_j. By Equation 4.49, we can define the individual injury probabilities

of the rule set in terms of the injury atoms. Applying Equations 4.48, 4.49 and 4.30 with Equation 4.19, we have

$$A_{i,j}(m,m) = 1 - \frac{\gamma_{i,j}(m)}{\gamma_j} A_j, \ m = 1, \cdots, 4 \tag{4.50}$$

Applying Equations 4.49 and 4.50 with Equation 4.20, we have

$$\pi_{i,j}(m) = \sum_{k=1}^{m-1} \pi_{i,j-1}(k) \cdot z_{i,j}(k,m) \cdot A_j + \pi_{i,j-1}(m) \cdot \left(1 - \frac{\gamma_{i,j}(m)}{\gamma_j} A_j \right) \tag{4.51}$$

By Equation 4.51, the individual injury probabilities of the rule set are defined on the injury atoms. The dependence among the rule sets can be modelled automatically through the injury atoms. We continue to discuss the order relationships among the rule sets.

4.6 Order Relationships among Rule Sets

Suppose that there are two rule sets $R_{i(1)}$ and $R_{i(2)}$. We call that the hazards associated with $R_{i(1)}$ are generally worse than those associated with $R_{i(2)}$ when

$$\gamma_{i(1),j}(1) \geq \gamma_{i(2),j}(1), \ j = 1, \cdots, D. \tag{4.52}$$

As defined in Equations 4.26 and 4.30, $\gamma_{i(1),j}(1)$ and $\gamma_{i(2),j}(1)$ are the means of $A_{i(1),j}(1)$ and $A_{i(2),j}(1)$ that are the general transition probabilities. By Equation 4.51, we have

$$\pi_{i,j}(1) = \pi_{i,j-1}(1) \cdot \left(1 - \frac{\gamma_{i,j}(1)}{\gamma_j} A_j \right)$$

Recalling $\pi_{i,0}(1)=1$, we have

$$\pi_i(1)=\pi_{i,D}(1)=\prod_{j=1}^{D}\left(1-\frac{\gamma_{i,j}(1)}{\gamma_j}A_j\right) \tag{4.53}$$

Based on the inequality 4.52 and recalling $0\le\gamma_j\le 1,\ 0\le A_j\le 1$, we have

$$1\ge\frac{\gamma_{i_1,j}(1)}{\gamma_j}A_j\ge\frac{\gamma_{i_2,j}(1)}{\gamma_j}A_j\ge 0$$

We then have

$$0\le 1-\frac{\gamma_{i_1,j}(1)}{\gamma_j(1)}A_j\le 1-\frac{\gamma_{i_2,j}(1)}{\gamma_j(1)}A_j\le 1,\ \ j=1,\cdots,D$$

It follows

$$\prod_{j=1}^{D}\left(1-\frac{\gamma_{i_1,j}(1)}{\gamma_j(1)}A_j\right)\le\prod_{j=1}^{D}\left(1-\frac{\gamma_{i_2,j}(1)}{\gamma_j(1)}A_j\right)$$

By Equation 4.53, we then have

$$\pi_{i(1)}(1)<\pi_{i(2)}(1) \tag{4.54}$$

By Equation 4.1, we have

$$\sum_{m=2}^{4}\pi_i(m)=1-\pi_i(1) \tag{4.55}$$

Applying Equation 4.55 with the inequality 4.54, we have

$$\sum_{m=2}^{4}\pi_{i(1)}(m)=1-\pi_{i(1)}(1)\geq 1-\pi_{i(2)}(1)=\sum_{m=2}^{4}\pi_{i(2)}(m) \tag{4.56}$$

We call $\sum_{m=2}^{4}\pi_{i(1)}(m)$ and $\sum_{m=2}^{4}\pi_{i(2)}(m)$ the general injury probabilities of the rule set $R_{i(1)}$ and $R_{i(2)}$ respectively. The inequality 4.56 shows that, when $R_{i(1)}$ has generally worse hazards than $R_{i(2)}$ does as defined in the inequality 4.52, the general individual injury probability of $R_{i(1)}$ is not less than that of $R_{i(2)}$. It is called the order relationship between $R_{i(1)}$ and $R_{i(2)}$.

4.7 A Simplified Model of Two Injury Levels

For uncertainty analysis, we always need to keep the elicitation workload reasonable for the experts. Usually we are more concerned with the major injuries and the fatalities than with the minor injuries. Therefore the minor injuries can be excluded in the uncertainty analysis. To reduce the elicitation work further, the major injuries and the fatalities can also be put into one category called casualties.

Accordingly, the model of the individual injury probabilities defined in Equation 4.51 can be simplified to accommodate only two levels of injuries: no injuries and casualties. We can exclude the minor injuries and the major injuries from the model defined in Equation 4.51 by setting

$$z_{i,j}(1,2)=z_{i,j}(1,3)=0,\ \ j=1,\cdots,D \tag{4.57}$$

By Equations 4.21 and 4.51, we then have

$$\pi_{i,j}(2)=\pi_{i,j}(3)=0,\ \ j=1,\cdots,D \tag{4.58}$$

Therefore by Equation 4.22, we have

$$\pi_i(2) = \pi_i(3) = 0 \tag{4.59}$$

Consequently the minor injuries and the major injuries are excluded from the model. We then denote by $\pi_{i,j}(4)$ the individual casualty probability. Applying Equation 4.57 with Equation 4.53, we have

$$\pi_i(1) = \prod_{j=1}^{D}\left(1 - \frac{z_{i,j}(1,4)}{\gamma_j} A_j\right) \tag{4.60}$$

By Equation 4.4, we have

$$\pi_i(4) = 1 - \pi_i(1) \tag{4.61}$$

Applying Equation 4.60 with Equation 4.61, we have

$$\pi_i(4) = 1 - \prod_{j=1}^{D}\left(1 - \frac{z_{i,j}(1,4)}{\gamma_j} A_j\right) \tag{4.62}$$

The Equation 4.62 defines a simplified model including only two levels of injuries: no injury and casualty. Such a model requires much less elicitation workload than the model of four injury levels. It is especially suitable when our concern is on the uncertainty of the major injuries and the fatalities. It has been used in some applications such as RSSB-SRM [Dennis 2006].

4.8 Distribution of an Injury Atom

As developed before, the injury atom measure the transition probability corresponding to the worst safety condition. Usually we assume that a random variable measuring the probability follows a Beta distribution on the interval [0,1]. Accordingly we assume

$$A_j \sim Beta(\alpha_j, \beta_j),\ A_j \in [0,1] \tag{4.63}$$

where α_j and β_j are the definition parameters of the Beta distribution.

The distribution of the injury atom can be built through expert judgement elicitation [O'Hagan 1998; O'Hagan, Buck et al. 2006]. We assume further that the injury atom has a unimodal probability distribution. Accordingly we impose the constraint that the definition parameters α and β are greater than 1 [Evans, Nicholas et al. 2000]. Consequently, the density probability and the cumulative probability of the injury atom A_j can be defined as

$$f(a_j;\alpha_j,\beta_j) = \frac{1}{B(\alpha_j,\beta_j)} \cdot a_j^{\alpha_j-1} \cdot (1-a_j)^{\beta_j-1} \tag{4.64}$$

and

$$F(a_j;\alpha_j,\beta_j) = \frac{B_{a_j}(\alpha_j,\beta_j)}{B(\alpha_j,\beta_j)} \tag{4.65}$$

where $a_j \in [0,1]$; $\alpha_j > 1$; $\beta_j > 1$; $B(\alpha_j,\beta_j)$ is a Beta function; $B_{a_j}(\alpha_j,\beta_j)$ is the incomplete Beta function.

We continue to develop a method for deriving α_j and β_j through the expert judgement elicitation. Based on Equations 4.64 and 4.65, the parameters α_j and β_j can be derived through the elicitation regarding A_j [O'Hagan, Buck et al. 2006]. By Equation 4.64, we have

$$E(A_j) = \frac{\alpha_j}{\alpha_j + \beta_j} \tag{4.66}$$

Recall that

$$E(A_j) = \gamma_j$$

We consequently have

$$\beta_j = \frac{1-\gamma_j}{\gamma_j}\alpha_j \tag{4.67}$$

For $P \in (0,1)$, we denote the $100 \times P$ percentile of A_j as $a_{j,P}$. Suppose the percentile $a_{j,P}$ has been elicited. By Equation 4.65, we have

$$F(A_j = a_{j,P}; \alpha_j, \beta_j) = a_{j,P} \tag{4.68}$$

By applying Equations 4.65 and 4.67 with Equation 4.68, an equation of α_j can be derived. Numerical algorithm can then be employed to solve α_j through the software package such as MATLAB or MS Excel. The parameter β_j can then be solved straightforwardly by Equation 4.67.

4.9 A Demonstration Example

4.9.1 Rule Sets and Injury Atoms

To demonstrate the methods for modelling the uncertainty in the rule sets through the injury atoms, an example is made on the risk assessment model RSSB-SRM [Dennis 2006]. In this model, the major injuries and the fatalities are put into one category called casualty. At the first stage of the uncertainty assessment, the minor injuries are concerned as much as the major injuries and the fatalities. Therefore, the model of two levels of injuries is employed for the uncertainty assessment.

In the demonstration example, we use three rule sets in RSSB-SRM including: T10-LCPRO-3, T10-LCPRO-4 and T10-LCPRO-7. The three rule sets are defined as:

T10-LCPRO-3: Train strikes large road vehicle above buffer height (low loader, JCB, etc.) on AHB, FP, MB/MCB/CCTV & all UWC (incl. MWL + T & UWC), no fire

T10-LCPRO-4: Train strikes large road vehicle above buffer height (low loader, JCB, etc.) on AHB, FP, MB/MCB/CCTV & all UWC (incl. MWL + T & UWC), with fire

T10-LCPRO-7: Train strikes HGV carrying flammable hazardous goods on AHB, FP, MB/MCB/CCTV & all UWC (incl. MWL + T & UWC) crossing, with fire

In the definitions of the three rule sets, the acronyms AHB, FP, MB/MCB/CCTV, and UWC stand for the different types of level crossings that are the outcomes of one escalation event [RSSB 2004; Dennis 2006]. From these three rule sets, we can identify two hazard types including: the mechanical impact and the fire. As discussed previously, the hazard type of mechanical impact is affected by the train's approach speeds that are associated with the types of the level crossings. A similar approach speed however is set for the level crossings included in the above three rule sets. Therefore the three rule sets have the same hazard associated with the mechanical impact. Depending on whether there are extra flammable goods carried in the involved road vehicle, two hazards associated with the fire are defined. The three hazards are summarised in Table 4.2. For each hazard, the mean of the casualty probability is elicited from experts and included in Table 4.2 as well.

Table 4.2 Hazards at different levels associated with the demonstration example

Code	Hazards	Mean of the casualty probability
MIPRO	Mechanical impact with the train approach speed on the other non-automotive level crossings	7.543%

Fire	Fire without extra flammable goods	2.857%
FireFGs	Fire with extra flammable goods carried in the involved road vehicle	14.2857%

4.9.2 Definition of Injury Atoms

As summarized in Table 4.2, one hazard associated with mechanical impact is defined, which is defined as the associated injury atom accordingly. Two hazards associated with fire are defined corresponding to whether there are extra flammable goods or not respectively. Between the two hazards associated with fire, the one with extra flammable goods represents the worse case and therefore is defined as the associated injury atom.

Within the RSSB-SRM, the upper bounds of the two injury atoms have already been elicited in addition to the means. As agreed with the experts of the RSSB-SRM, the upper bound is interpreted as the 99 percentile. The means and the 99 percentiles of the two injury atoms are summarized in Table 4.3. We assume the injury atoms follow Beta distributions. The definition parameters α and β of the two injury atoms can then be calculated by the above procedure and are summarized in Table 4.3.

Table 4.3 Definitions of the injury atoms for the demonstration example

Code	The Mean	99 percentile	α	β	Std
MIPRO	7.543%	15.09%	7.09	86.93	0.027
FireFGs	14.28%	28.57%	6.27	37.62	0.052

4.9.3 Definition of Rule Sets on Injury Atoms

The three rule sets in the example can then be defined on the two injury atoms by Equation 4.49. The weights associated with the injury atoms are calculated by Equation 4.48 and are summarized in Table 4.4. The means of the rule sets are calculated by Equation 4.68 and are included in Table 4.4 as well.

Table 4.4 Definitions of the rule sets on the injury atoms for the demonstration example

Rule Sets	MIPRO	FireFGs	The mean
T10-LCPRO-3	1	0	0.07543
T10-LCPRO-4	1	0.2	0.10185
T10-LCPRO-7	0	1	0.14286

With the injury atoms defined, we can continue to investigate the dependence and the order relations among the rule sets. For this purpose, we employ the Monte-Carlo simulations. At first, we generate independently the samples of the two injury atoms by the definitions as summarised in Table 4.3. The individual casualty probability of the three rule sets are then calculated by Equation 4.62 based on the weights summarized in Table 4.4. Based on 6000 samples of the three rule sets, the scatter plots are made pairwise to demonstrate the dependence and order relationship among the three rule sets. Demonstrated in Fig. 4.4 is the scatter plot between the rule sets T10-LCPPRO-4 and T10-LCPPRO-7. By Table 4.4, the two rule sets share the injury atom FireFGs. For T10-LCPPRO-4, however, the injury atom MIPRO is more than twice stronger than the injury atom FireFGs in terms of the mean of the individual casualty probability. Consequently a moderate dependence is introduced for the two rule sets as shown in Fig. 4.4. There is no order relation demonstrated between the rule sets T10-LCPPRO-4 and T10-LCPPRO-7.

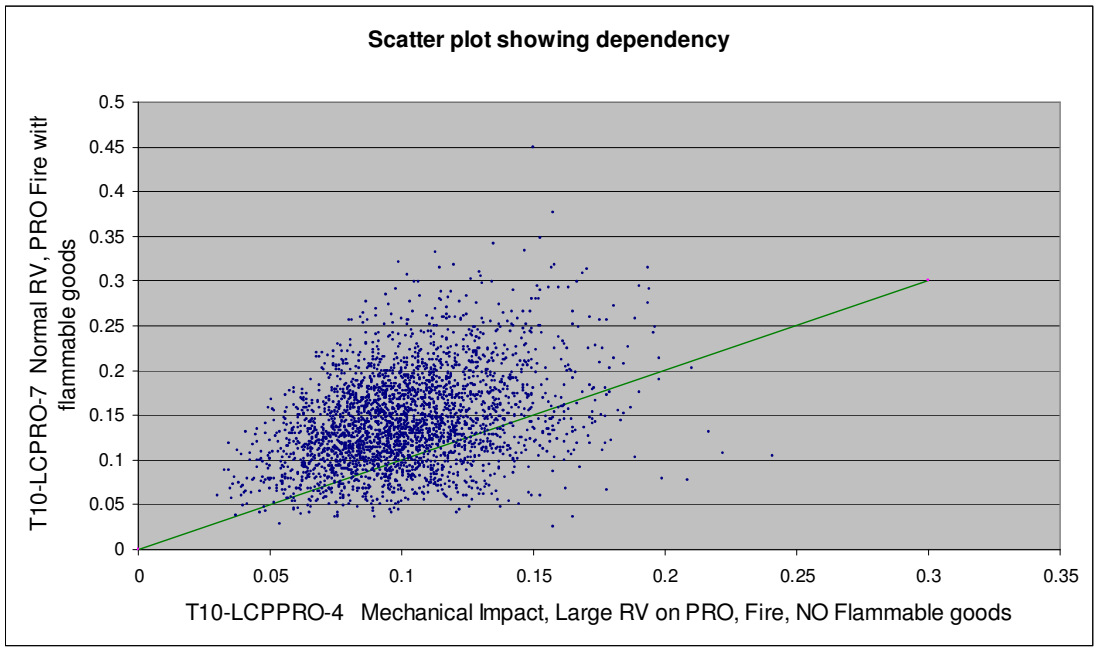


Fig. 4.4 Scatter plot between T10-LCPPRO-4 and T10-LCPPRO-7

Demonstrated in Fig. 4.5 is the scatter plot between T10-LCPPRO-4 and T10-LCPPRO-3. As summarized in Table 4.4 the two rule sets share the injury atom MIPRO. The rule set T10-LCPPRO-4 has an extra injury atom FireFGs. Because the injury atom MIPRO is more than twice stronger than the injury atom FireFGs in terms of the mean of the individual casualty probability, the rule sets T10-LCPPRO-4 and T10-LCPPRO-3 have a stronger common injury atom than T10-LCPPRO-4 and T10-LCPPRO-7 do. As a result, a stronger dependence is induced between T10-LCPPRO-4 and T10-LCPPRO-3 than that between T10-LCPPRO-4 and T10-LCPPRO-7 as shown in Fig. 4.5. Due to the extra injury atom FireFGs, T10-LCPPRO-4 should have an individual casualty probability always larger than T10-LCPPRO-3 does. This order relationship is also demonstrated in Fig. 4.5 where the all sample points fall under the 45-degree line.

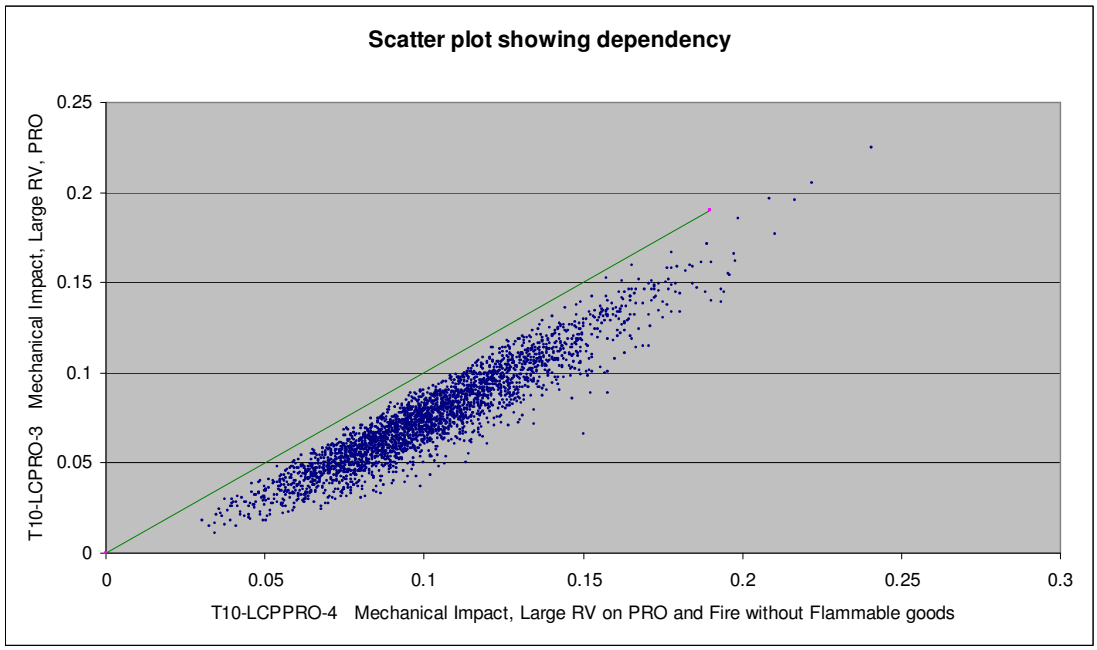


Fig. 4.5 Scatter plot between T10-LCPPRO-4 and T10-LCPPRO-3

4.10 Conclusion

In this chapter, we have developed a method for modelling the uncertainty in the rule sets of an event tree. Usually the experts make their assessments of consequences of risk scenarios by thinking of the included hazards separately and then aggregating them together, as is the case with RSSB-SRM. We try to construct a model of the correlations between the parameters and reconstruct the reasoning implicit behind the calculation.

Along the sequence of the escalation events of an event tree, we identify the hazard types that are the materials or activities with potential to cause injuries. Depending on the outcomes of the escalation events associated with a risk scenario, each hazard type is defined at a specific level, which is called a hazard. The individual injury probabilities of the rule set are then defined in terms of the hazards.

We assume that, given the sequence, the hazards are independent in terms of causing the injuries to the people exposed to the risk scenario. When an individual person

caught in a risk scenario is not lucky enough to escape all the hazards, the person becomes part of the consequence. After each hazard in the sequence, an individual exposed to the risk scenario either keeps the same level of injury or suffers a higher level injury. We assume that the transition probabilities associated with a hazard depend only on the person's injury level before the hazard. Consequently, the individual injury probabilities of a rule set can be defined on the associated hazards by a Markov Chain model [Ross 2003]. To build up the Markov Chain model, we need to elicit the transition probabilities associated with each hazard. Once the transition probabilities are elicited for one hazard, they can be used for all the rule sets containing the same hazard. Therefore, modelling on the hazards can significantly reduce the elicitation work.

For modelling the uncertainty in the rule sets, we need to model the uncertainty in the transition probabilities that belongs to epistemic uncertainty category as well. We assume that the transition probabilities associated with different hazard types to be independent. It means that a given transition probability of one hazard does not tell any information on the transition probability of the hazards of another type. We assume that the transition probabilities associated with the same hazard type are proportional. By the assumption, an injury atom is defined for each hazard type. All the transition probabilities associated with the same hazard type are then defined on the injury atom. Consequently the individual injury probabilities of the rule sets are modelled on the injury atoms.

In most of the cases, we are more concerned with the major injuries and the fatalities than with the minor injuries. We need also to keep the elicitation work and time reasonable to the experts. For these two purposes, we develop a model of two levels of injuries as a simplified case of the model of four levels of injuries.

By the above methods, we elicit the uncertainties regarding the hazards; the uncertainty in the rule sets are then defined in terms of the hazards. The number of the hazards is much smaller than that of the rule sets. Therefore, the above methods require reasonable elicitation time from experts. By the above method, the

dependences among the rule sets are modelled automatically through the injury atoms. The monotonicity property of the rule sets can also be kept between two rule sets associated with general worse and better hazards respectively.

For future research work, we can investigate the possible application of the ordered Dirichlet distribution on modelling transition probabilities.

Chapter 5

Uncertainty Assessment of Fault-tree and Event-tree Models

5.1 Overview

Fault trees and event trees are widely used in probabilistic risk analysis. The uncertainty in the output of the Fault-tree and Event-tree model is driven by: (1) the uncertainty in the basic events of the fault tree; (2) the uncertainty in the probabilities of the outcomes of the escalation events; and (3) the uncertainty in the consequences of the risk scenarios. In this research, we confine the Fault-tree and Event-tree model as follows. The subjective uncertainty of the basic events is assumed to follow a joint normal distribution. The probabilities of the escalation events' outcomes are set as the best-estimated values; the uncertainty in these probabilities is not studied in this research. The consequence of a risk scenario is defined as the means of equivalent fatalities counting both the fatalities and the scaled-down major injuries, which is the product of the number of the people exposed to the risk scenario and the individual casualty probability of the associated rule set. The uncertainty in the rule set is modelled on the injury atoms by the simplified model of two injury levels as developed in Chapter 4. The numbers of people exposed to the risk scenarios are set with the best-estimated values; the uncertainty in these numbers is not studied in this research. Focusing on the fault trees and event trees as described above, we are going to develop the methods for conducting the subjective uncertainty analysis.

At the first step, we develop a mimic model of the fault tree and event tree model. Fault trees and event trees are usually built with commercial software packages such

as Isograph FT+. The software packages usually offer a graphic interface for the users to build up the models. The software then transfers the graphic model into computer codes and data that are stored in the internal database. After a run of the model, the generated results are also stored in the internal database. The internal databases, however, are usually not transparent to the users. To access the internal database, users must utilize the dedicated interface software tools that are developed associated with the commercial software packages. Associated with Isograph FT+, for example, we need to program with the dynamical link library (DLL) to access the internal database and to run the models [Isograph 2004; Isograph 2005]. These limits on accessing and manipulating the internal database cause difficulties in conducting uncertainty assessment in two aspects. Firstly, for conducting simulations, we need to set the input parameters, run the model and obtain the results. It is very difficult if possible given the limits on accessing the database of the fault tree and event tree computer model. Secondly, the computer model consisting of the codes and data stored in the internal database is completely a "black-box" to the analysts. It is impossible to do any analytical analysis with such a computer model. Therefore we need a mimic model of the original computer model.

The top event of a fault tree is broken down into the basic events, also called precursors. A cut set is a collection of the basic events that together certainly cause the top event. A minimum cut set is one that is no longer a cut set when any of its basic events is removed. The basic events are called rare events when they have very small occurrence probabilities. For two minimum cut sets composed of rare events, the simultaneous occurrence probability is an order of magnitude smaller than the occurrence probability of either minimum cut sets. Based on this idea, the occurrence probability of the top event can be approximated as the sum of the occurrence probabilities of all the minimum cut sets, which is called rare event approximation [Bedford and Cooke 2001]. Once the fault trees and event trees are built with a commercial software tool such as Isograph and the input parameters are set, the minimum cut sets can be generated and associated with a risk scenario. The codes of the events composing the risk scenarios can then be output into a plain text file such as MS Excel. We can then program with MS Excel VBA on the events composing

the risk scenarios to mimic the original computer model [Jelen and Syrstad 2004]. MS Excel is a very popular software tool that offers easy access to the data. It is therefore easy to conduct simulations of the mimic model for uncertainty assessment. The mimic model also offers a transparent structure so that we can calculate the variance of the mimic model through algebraic operations.

For conducting the uncertainty analysis, the dependent basic events need to be expressed as the linear transformation of a set of independent standard normal variables. We select a linear transformation that is suitable for the context of risk analysis. Usually we implement the linear transformation through the *Cholesky* decomposition of the covariance matrix [Scheuer and Stoller 1962]. The *Cholesky* algorithm requires that the covariance matrix must be positive definite [Scheuer and Stoller 1962]. In the context of risk analysis, however, the covariance matrix can be positive semi-definite. To solve this issue, we select the linear transformation through the decomposition vectors and eigenvalues, which is called eigen-decomposition. Furthermore, in the context of risk analysis, the variance of the input parameters can be very small. The variance of the precursors related to the hazardous event HET-12 in the RSSB-SRM, for example, spreads over $10^{-19} \sim 10^{-24}$ [Harrison, Griffin et al. 2008]. For such a covariance matrix, the calculation of the eigenvalues and eigenvectors is more expensive and large errors can be incurred [Wilkinson 1965; Watkins 1991]. To solve this problem, we implement the linear transformation through the eigen-decomposition of the correlation matrix instead of the covariance matrix.

Based on the linear transformation, we then develop Monte-Carlo simulations to build the empirical distribution of the output. We also develop the analytical solution for the variance through algebraic operations. The two methods are implemented independently and therefore can be used for cross check for each other. The whole scheme of the uncertainty assessment of a Fault-tree and Event-tree model is demonstrated in Fig. 5.1. Blocks 1, 2 and 3 represent the procedures for building the Excel mimic model, which are to be developed in Section 5.2. Block 5 represents the procedure for building the uncertainty model in the precursors through elicitation,

which is related to Chapter 3. Block 7 represents the procedure for building the uncertainty model in the rule sets, which is related to Chapter 4. The uncertainty model in the precursors and the uncertainty model in rule sets are represented in Block 4 and Block 6 respectively, as the outcomes of Block 5 and Block 7. Block 8 and 9 represent the procedures for uncertainty assessment, for which a set of methods are to be developed in this chapter. The software design for implementing the mimic model, the simulations and the analytical solution of the variance is developed in Section 5.7.

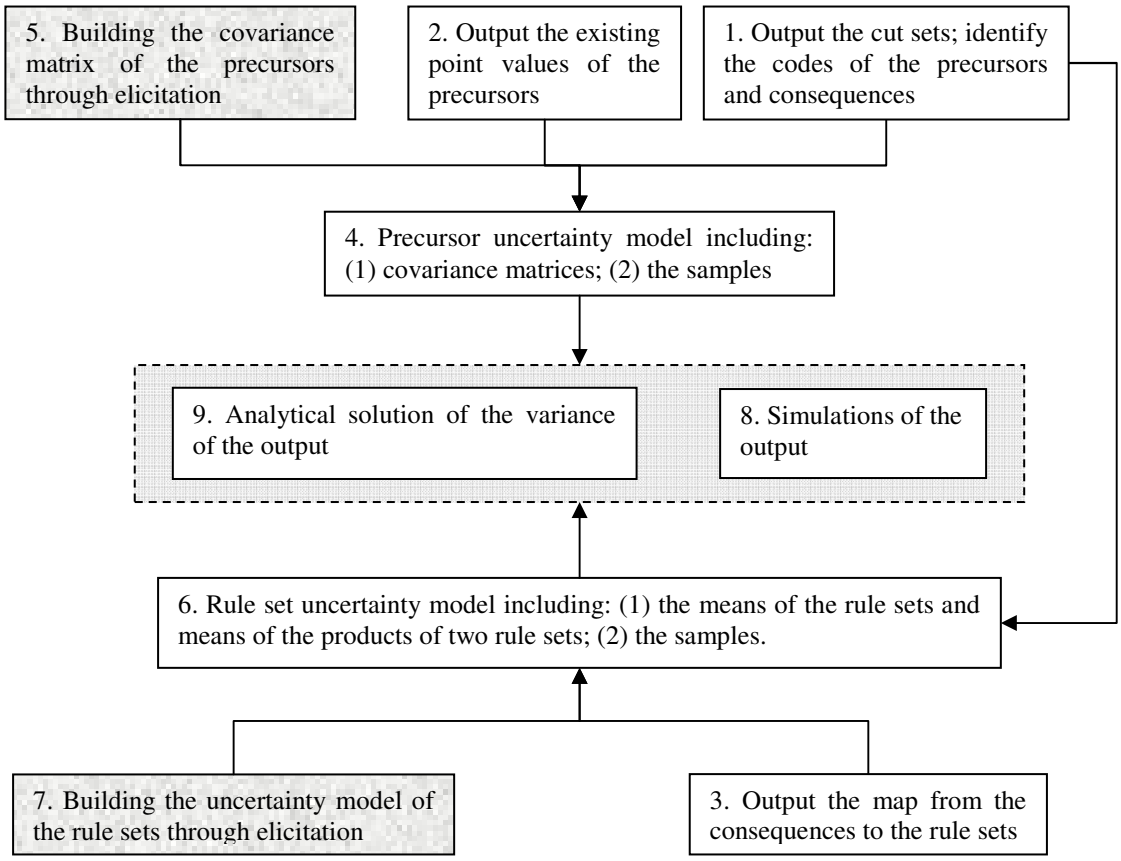


Fig. 5.1 Scheme for conducting the uncertainty assessment of a Fault-tree and Event-tree model

In Appendix A, a set of new methods are also proposed. These methods need to be developed further especially with techniques for error control. Consequently they are

not applied in this research. These methods, however, have the potential to calculate the uncertainty propagation and are worth more effort in the future.

5.2 Mimic Model with MS Excel

Usually we use an event tree to model the risk scenarios. A path from the initiating event through all the escalation events is called an accident sequence. At the end of an accident sequence, the risk scenario is defined. The initiating event of the event tree can be further broken down into the basic events with a fault tree. A cut set is a collection of the basic events that together certainly cause the top event of the fault tree. A minimum cut set is one that is no longer a cut set when any of its basic events is removed. When the rare event approximation is applied, the occurrence of the top event can be approximated by the sum of the occurrences of the minimum cut sets [Bedford and Cooke 2001]. Consequently a finite set of risk scenarios can be defined by the combinations of the minimum cut sets and the outcomes of the events along the event tree.

The frequency of the risk scenario S_ℓ can be calculated as

$$f_\ell = \prod_{p_i \in S_\ell} p_i \cdot \prod_{e(\ell,i) \in S_\ell} \Pr(e(\ell,i)) \tag{5.1}$$

where p_i stands for a basic event; $e(\ell,i)$ stands for the outcome of the escalation event E_i associated with S_ℓ; $\Pr(e(\ell,i))$ stands for the probability of $e(\ell,i)$.

Suppose that the consequence associated with S_ℓ is c_ℓ, the risk associated with S_ℓ is calculated as

$$r_\ell = f_\ell \cdot c_\ell \tag{5.2}$$

The overall risk associated with the fault-tree and event tree can then be defined as

$$r = \sum_{\ell=1}^{N(S)} f_\ell \cdot c_\ell \tag{5.3}$$

where $N(S)$ is the number of risk scenarios of the fault tree and event tree.

Using Equations 5.1-5.3, we can define a mimic model of the fault tree and event tree when we obtain for each risk scenario: (1) the codes of the basic events; (2) the codes of the outcomes of the escalation events; and (3) the code of the consequence. These codes are referred to as the risk scenario definition codes. Once the fault tree and event tree are built with the commercial software, the risk scenario definition codes can be generated and then output in plain text or MS Excel etc. The software Isograph FT+, for example, can output the risk scenario definition codes into an Excel worksheet. An example of this is demonstrated in Fig. 5.2.a. On the top row are the titles of the fields as summarized in Table 5.1; underneath the top row, each row defines a risk scenario. While most fields are output for explanation or cross-checking, the fields *Cut Set* and *Consequence Name* are the two key fields containing the risk scenario definition codes.

The field *"Cut Set"* consists of a character string that is composed of the codes of the basic events and the codes of the outcomes of the escalation events defining the risk scenario [Isograph 2005]. An example of the Cut Set string is extracted as

"OPEN-TRACK. EL-N----PI. ELCDAOCSPP. -ELFS----PP. PABCLOVSTH"*.

The character string can be parsed by programming into the individual codes. These codes can then the stored separately in the following columns as shown in Fig. 5.2.b. By the codes of the events and the outcomes, we can link the mimic model to the uncertainty model of the input parameters.

The filed *"Consequence Name"* contains the code of the consequence. As explained in Chapter 4, a consequence is associated with a rule set and the number of the people exposed to the risk scenario. The rule set defines the individual injury

probabilities, of which the uncertainty model can be built on the injury atoms as developed in Chapter 4. Therefore by the code of the consequence, we can link the mimic model to the uncertainty model of the rule sets.

As a conclusion, we can mimic the Fault-tree and Event-tree model by Equations 5.1-5.3 with the risk scenario definition codes. MS Excel is a very popular software tool and we can easily manipulate the data within Excel. With the mimic model, we can conveniently conduct simulations and calculate the analytical solution of the variance of the output risk. We will continue to develop these methods by starting with the linear transformation of the normal random variables.

Table 5.1 Summary of the fields defining the risk scenarios in the mimic model

No	Code	Description
1.	Risk	The risk associated with the associated risk scenario
2.	Frequency	The frequency of the risk scenario
3.	Weight	The consequence, i.e. the number of the fatalities associated with the risk scenario
4.	Cut Set	Character string composing the codes of the basic events and the codes of the outcomes of the escalation events
5.	ID	A unique index code of the risk scenario
6.	Event Description	Description regarding to the risk scenario
8.	Consequence Name	Index code of the consequence associated with the risk scenario
9.	Consequence Description	Description of the consequence

Risk	Frequency	Weight	Cut Set	ID	Event Desc	Fussell-Ve	Consequence	Name	Consequer
2.3E-05	0.003258	0.007059	OPEN-TRA	T12-OFN-1	Number of	3.42E-05	T12-OFN-1		OPEN-TRA
6.09E-06	0.000862	0.007059	OPEN-TRA	T12-OFN-1	Number of	9.05E-06	T12-OFN-1		0.007 (10)
6.76E-07	9.58E-05	0.007059	OPEN-TRA	T12-OFN-1	Number of	1.01E-06	T12-OFN-1		0.007 (10)
4.17E-07	5.9E-05	0.007059	OPEN-TRA	T12-OFN-1	Number of	6.2E-07	T12-OFN-1		0.007 (10)
3.29E-07	4.66E-05	0.007059	OPEN-TRA	T12-OFN-1	Number of	4.89E-07	T12-OFN-1		0.007 (10)
1.86E-07	2.63E-05	0.007059	OPEN-TRA	T12-OFN-1	Number of	2.77E-07	T12-OFN-1		0.007 (10)
1.75E-07	2.48E-05	0.007059	OPEN-TRA	T12-OFN-1	Number of	2.6E-07	T12-OFN-1		0.007 (10)
1.55E-07	2.19E-05	0.007059	OPEN-TRA	T12-OFN-1	Number of	2.3E-07	T12-OFN-1		0.007 (10)
1.35E-07	1.92E-05	0.007059	OPEN-TRA	T12-OFN-1	Number of	2.01E-07	T12-OFN-1		0.007 (10)
1.33E-07	1.89E-05	0.007059	OPEN-TRA	T12-OFN-1	Number of	1.99E-07	T12-OFN-1		0.007 (10)
5.68E-08	8.05E-06	0.007059	OPEN-TRA	T12-OFN-1	Number of	8.45E-08	T12-OFN-1		0.007 (10)
2.11E-09	2.99E-07	0.007059	OPEN-TRA	T12-OFN-1	Number of	3.14E-09	T12-OFN-1		0.007 (10)
2.11E-09	2.99E-07	0.007059	OPEN-TRA	T12-OFN-1	Number of	3.14E-09	T12-OFN-1		0.007 (10)
2.03E-09	2.88E-07	0.007059	OPEN-TRA	T12-OFN-1	Number of	3.03E-09	T12-OFN-1		0.007 (10)
2.03E-09	2.88E-07	0.007059	OPEN-TRA	T12-OFN-1	Number of	3.03E-09	T12-OFN-1		0.007 (10)
1.64E-09	2.32E-07	0.007059	OPEN-TRA	T12-OFN-1	Number of	2.44E-09	T12-OFN-1		0.007 (10)
1.58E-09	2.24E-07	0.007059	OPEN-TRA	T12-OFN-1	Number of	2.35E-09	T12-OFN-1		0.007 (10)
2.11E-10	2.99E-08	0.007059	OPEN-TRA	T12-OFN-1	Number of	3.14E-10	T12-OFN-1		0.007 (10)
2.04E-10	2.89E-08	0.007059	OPEN-TRA	T12-OFN-1	Number of	3.03E-10	T12-OFN-1		0.007 (10)

Fig. 5.2.a MS Excel mimic model of the Isograph FT+ Fault-tree and Event-tree models

Number Events	C1	C2	C3	C4	C5	C6
5	OPEN-TRACK	EL-N----PI	ELCDAOCSPP	-ELFS---PP	VAOCLDRRTH	
5	OPEN-TRACK	EL-N----PI	ELCDAOCSPP	-ELFS---PP	VAOCLDELTH	
5	OPEN-TRACK	EL-N----PI	ELCDAOCSPP	-ELFS---PP	VAOCLDRTTH	
5	OPEN-TRACK	EL-N----PI	ELCDAOCSPP	-ELFS---PP	VABCLDRRTH	
5	OPEN-TRACK	EL-N----PI	ELCDAOCSPP	-ELFS---PP	VAOCLSUIXPE	
5	OPEN-TRACK	EL-N----PI	ELCDAOCSPP	-ELFS---PP	WAOCLENVTE	
5	OPEN-TRACK	EL-N----PI	ELCDAOCSPP	-ELFS---PP	VABCLSUIXPE	
5	OPEN-TRACK	EL-N----PI	ELCDAOCSPP	-ELFS---PP	VABCLDELTH	
5	OPEN-TRACK	EL-N----PI	ELCDAOCSPP	-ELFS---PP	VAOCLSTRTE	
5	OPEN-TRACK	EL-N----PI	ELCDAOCSPP	-ELFS---PP	VABCLSTRTE	
5	OPEN-TRACK	EL-N----PI	ELCDAOCSPP	-ELFS---PP	WABCLENVTE	
5	OPEN-TRACK	EL-N----PI	ELCDAOCSPP	-ELFS---PP	LAOCLLSETF	
5	OPEN-TRACK	EL-N----PI	ELCDAOCSPP	-ELFS---PP	PAOCLOVSTH	
5	OPEN-TRACK	EL-N----PI	ELCDAOCSPP	-ELFS---PP	PABCLOVSTH	
5	OPEN-TRACK	EL-N----PI	ELCDAOCSPP	-ELFS---PP	LABCLBLETF	
6	OPEN-TRACK	EL-N----PI	ELCDAOCSPP	-ELFS---PP	PAOCLSPDTH	PSPDDRECPH
6	OPEN-TRACK	EL-N----PI	ELCDAOCSPP	-ELFS---PP	PABCLSPDTH	PSPDDRECPH
6	OPEN-TRACK	EL-N----PI	ELCDAOCSPP	-ELFS---PP	PAOCLSPDTH	PSPDMISCPH
6	OPEN-TRACK	EL-N----PI	ELCDAOCSPP	-ELFS---PP	PABCLSPDTH	PSPDMISCPH

Fig. 5.2.b Codes of the events composing the risk scenario.

5.3 Linear Transformation of Dependent Normal Random Variables

5.3.1 Overview

Suppose a column vector of N normal random variables is defined as

$$\mathbf{p} = [p_1, \cdots, p_N]^T \tag{5.4}$$

We define the means and the covariance matrix of $\mathbf{p}$ as

$$E(\mathbf{p}) = [E(p_1), \cdots, E(p_N)]^T \tag{5.5}$$

$$\text{cov}(\mathbf{p}) = \mathbf{\Sigma_p}, \ \mathbf{\Sigma_p} \in \mathbb{R}^{N \times N} \tag{5.6}$$

The joint distribution of $\mathbf{p}$ is then completely defined by $E(\mathbf{p})$ and $\text{cov}(\mathbf{p})$. It is however difficult to sample or to conduct other calculation directly on the dependent normal random variables. Therefore we prefer to define $\mathbf{p}$ in terms of a linear transformation of a set of independent normal random variables. Suppose there are N independent standard normal random variables that are defined as

$$Z_i \sim N(0,1), \ i = 1, \cdots, N \tag{5.7}$$

$$\text{cov}(\mathbf{Z}) = \mathbf{I}_N \tag{5.8}$$

where $\mathbf{I}_N$ is an $N \times N$ dimensional identity matrix.

Accordingly we define the column vector

$$\mathbf{Z} = (Z_1, \cdots, Z_N)^T \tag{5.9}$$

We search for a linear transformation from $\mathbf{Z}$ to $\mathbf{p}$ of the form

$$\mathbf{p} = \mathbf{L} \cdot \mathbf{Z} + E(\mathbf{p}) \tag{5.10}$$

where $\mathbf{L} \in \mathbb{R}^{N \times N}$ stands for a linear transformation matrix.

We know that the joint normal distribution of $\mathbf{p}$ can be completely defined by the means $E(\mathbf{p})$ and the covariance matrix $\mathbf{\Sigma_p}$. The transformation matrix $\mathbf{L}$ can be derived in terms of $E(\mathbf{p})$ and $\mathbf{\Sigma_p}$. Based on Equation 5.7, we have

$$E(\mathbf{Z}) = \mathbf{0} \tag{5.11}$$

Consequently we have

$$E(\mathbf{L} \cdot \mathbf{Z}) = \mathbf{L} \cdot E(\mathbf{Z}) = \mathbf{0}$$

Therefore the mean vector $E(\mathbf{p})$ is automatically preserved by the linear transformation. By Equation 5.10, we have

$$\text{cov}(\mathbf{p}) = \text{cov}(\mathbf{L} \cdot \mathbf{Z}) = \mathbf{L} \cdot \text{cov}(\mathbf{Z}) \cdot \mathbf{L}^T$$

Based on Equations 5.6 and 5.8, we then have

$$\mathbf{\Sigma_p} = \mathbf{L}\mathbf{L}^T \tag{5.12}$$

Therefore once the transformation matrix $\mathbf{L}$ is built to satisfy Equation 5.12, the linear transformation in Equation 5.10 will preserve the covariance matrix $\mathbf{\Sigma_p}$ as required. We know that the covariance matrix $\mathbf{\Sigma_p}$ must be symmetrical and positive semi-definite. When $\mathbf{\Sigma_p}$ is positive definite, $\mathbf{L}$ can be solved through *Cholesky* decomposition [Scheuer and Stoller 1962; Herstein and Winter 1988; Law and McComas 1999], where the outcome $\mathbf{L}$ is a lower triangular matrix.

In risk analysis context, however, the covariance matrix $\mathbf{\Sigma_p}$ can be positive semi-definite. The *Cholesky* cannot be used in this situation [Scheuer and Stoller 1962; Law and McComas 1999]. To solve this problem, we are going to implement the

linear transformation through eigen-decomposition as used in principal component analysis [Krzanowski 1988; Jolliffe 2002].

5.3.2 Linear Transformation through Eigen-decomposition

We know that the covariance matrix $\mathbf{\Sigma_p}$ must be symmetric matrix and positive semi-definite. Therefore, associated with $\mathbf{\Sigma_p}$, there exist N non-negative real eigenvalues $\lambda_i \geq 0 \quad i = 1, \cdots, N$, [Burden and Faires 1997]. Corresponding to each $\lambda_i \geq 0$, there exists an eigenvector $\mathbf{U}_i \in \mathbb{R}^{N \times 1}$ that satisfies

$$\mathbf{\Sigma_p U}_i = \lambda_i \mathbf{U}_i, \; i = 1, \cdots, N \tag{5.13}$$

We define the eigenvalue matrix as

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_N \end{bmatrix} \tag{5.14}$$

Therefore $\mathbf{\Lambda}$ is a diagonal matrix with all the eigenvalues on the diagonal. According to $\mathbf{\Lambda}$, we define the eigenvector matrix as

$$\mathbf{U} = [\mathbf{U}_1, \cdots, \mathbf{U}_N], \; \mathbf{U} \in \mathbb{R}^{N \times N} \tag{5.15}$$

By Equations 5.13, 5.14 and 5.15, we have

$$\mathbf{\Sigma_p U} = \mathbf{U\Lambda} \tag{5.16}$$

Usually the eigenvectors are normalized so that

$$\mathbf{U}_i^T \cdot \mathbf{U}_i = 1 \tag{5.17}$$

The eigenvectors $\mathbf{U}_i$, $i = 1, \cdots, N$, are mutually orthogonal [Burden and Faires 1997]. Consequently we have

$$\mathbf{U}_i^T \cdot \mathbf{U}_j = 0, \; i \neq j \tag{5.18}$$

By Equations 5.17 and 5.18, we have

$$\mathbf{U}^T \cdot \mathbf{U} = \mathbf{I}_N \tag{5.19}$$

Therefore the eigenvector matrix $\mathbf{U}$ is orthogonal. By the definition of the inverse matrix, we have

$$\mathbf{U}^{-1} = \mathbf{U}^T \tag{5.20}$$

where $\mathbf{U}^{-1}$ stands for the inverse matrix of $\mathbf{U}$.

Therefore we have

$$\mathbf{U}\mathbf{U}^T = \mathbf{I}_N \tag{5.21}$$

where $\mathbf{I}_N$ stands for the $N \times N$ dimensional identity matrix as defined before.

Based on Equation 5.21, by right-multiplying $\mathbf{U}^T$ on both sides of Equation 5.16 we have

$$\mathbf{\Sigma_p} = \mathbf{\Sigma}\mathbf{U}\mathbf{U}^T = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T \tag{5.22}$$

From Equation 5.14, we define

$$\mathbf{\Lambda}^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{\lambda_N} \end{bmatrix} \tag{5.23}$$

Therefore we have

$$\mathbf{\Lambda} = \mathbf{\Lambda}^{\frac{1}{2}} \cdot \mathbf{\Lambda}^{\frac{1}{2}}, \quad \left(\mathbf{\Lambda}^{\frac{1}{2}} \right)^T = \mathbf{\Lambda}^{\frac{1}{2}} \tag{5.24}$$

Based on Equations 5.22, 5.23 and 5.24, we have

$$\mathbf{\Sigma_p} = \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}} \cdot \left(\mathbf{\Lambda}^{\frac{1}{2}} \right)^T \mathbf{U}^T = \left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}} \right) \left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}} \right)^T \tag{5.25}$$

Based on Equations 5.12 and 5.25, we can define

$$\mathbf{L} = \mathbf{U} \cdot \mathbf{\Lambda}^{\frac{1}{2}}$$

Since $\mathbf{L}$ is defined by the eigenvectors and the eigenvalues, the decomposition defined in Equation 5.25 is called eigen-decomposition [Dillon and Goldstein 1984]. Based on Equations 5.10, we can define the linear transformation as

$$\mathbf{p} = \left(\mathbf{U} \cdot \mathbf{\Lambda}^{\frac{1}{2}} \right) \cdot \mathbf{Z} + E(\mathbf{p}) \tag{5.26}$$

where $\mathbf{Z} = (Z_1, \cdots, Z_N)^T$ is a column vector of independent standard normal random variables as defined in Equations 5.7-5.9.

The linear transformation through Eigen-decomposition can be used when the covariance matrix $\mathbf{\Sigma_p}$ is positive semi-definite; while the linear transformation through the *Cholesky* decomposition can be used only when $\mathbf{\Sigma_p}$ is positive definite.

In the uncertainty analysis of risk assessment models, the variance of the input parameters can be very small. The variance of the precursors related to the hazardous event HET-12 in the RSSB-SRM, for example, spreads over $10^{-19} \sim 10^{-24}$ [Harrison, Griffin et al. 2008]. We have

$$\mathrm{tr}(\mathbf{\Sigma_p}) = \sum_{i=1}^{N} \sigma_i^2 = \sum_{i=1}^{N} \lambda_i \tag{5.27}$$

where $\mathrm{tr}(\mathbf{\Sigma_p})$ is the trace of $\mathbf{\Sigma_p}$; σ_i stands for the standard deviation of p_i.

Because $\mathbf{\Sigma_p}$ is positive semi-definite, we have $\lambda_i \geq 0$, $i = 1, \cdots, N$. By Equation 5.27, we have

$$0 \leq \lambda_i \leq \sum_{i=1}^{N} \sigma_i^2$$

When all the standard deviations are very small, the eigenvalues of $\mathbf{\Sigma_p}$ cluster in the small interval $\left[0, \sum_{i=1}^{N} \sigma_i^2\right]$. For such a covariance matrix $\mathbf{\Sigma_p}$, the calculation of the eigenvalues and eigenvectors is more expensive while large errors are incurred [Wilkinson 1965; Watkins 1991]. To solve this problem, we are going to develop the linear transformation through the eigen-decomposition of the correlation matrix.

5.3.3 Linear Transformation based on the Correlation Matrix

Suppose the correlation matrix of $\mathbf{p}$ is defined as

$$r(\mathbf{p}) = \begin{bmatrix} 1 & \cdots & r(X_1, X_i) & \cdots & r(X_1, X_N) \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ r(X_i, X_1) & \cdots & 1 & \cdots & r(X_i, X_N) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r(X_N, X_1) & \cdots & r(X_N, X_i) & \cdots & 1 \end{bmatrix}$$

The correlation matrix $r(\mathbf{p})$ is symmetric and positive semi-definite. Therefore $r(\mathbf{p})$ can be decomposed with the eigenvalue matrix and the eigenvector matrix as

$$r(\mathbf{p}) = \mathbf{V} \cdot \mathbf{M} \cdot \mathbf{V}^T = \left(\mathbf{V}\mathbf{M}^{\frac{1}{2}} \right) \cdot \left(\mathbf{V}\mathbf{M}^{\frac{1}{2}} \right)^T \tag{5.28}$$

where $\mathbf{V}$ and $\mathbf{M}$ stand for the eigenvector matrix and the eigenvalue matrix and of $r(\mathbf{p})$ respectively.

We define

$$\boldsymbol{\sigma}_\mathbf{p} = \begin{bmatrix} \sigma_1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \cdots & \vdots \\ 0 & \cdots & \sigma_i & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & \sigma_N \end{bmatrix}$$

where σ_i stands for the standard deviation of p_i.

Therefore $\boldsymbol{\sigma}_\mathbf{p}$ is a diagonal matrix and we have

$$\boldsymbol{\sigma}_\mathbf{p} = \boldsymbol{\sigma}_\mathbf{p}^T \tag{5.29}$$

Consequently we have

$$\boldsymbol{\Sigma}_{\mathbf{p}} = \boldsymbol{\sigma}_{\mathbf{p}} \cdot r(\mathbf{p}) \cdot \boldsymbol{\sigma}_{\mathbf{p}} \tag{5.30}$$

By applying Equation 5.28 with Equation 5.30 we have

$$\begin{aligned} \boldsymbol{\Sigma}_{\mathbf{p}} &= \boldsymbol{\sigma}_{\mathbf{p}} \cdot \left(\mathbf{V} \cdot \mathbf{M}^{\frac{1}{2}} \right) \cdot \left(\mathbf{V} \cdot \mathbf{M}^{\frac{1}{2}} \right)^{T} \cdot \boldsymbol{\sigma}_{\mathbf{p}} \\ &= \left(\boldsymbol{\sigma}_{\mathbf{p}} \cdot \mathbf{V} \cdot \mathbf{M}^{\frac{1}{2}} \right) \cdot \left(\boldsymbol{\sigma}_{\mathbf{p}} \cdot \mathbf{V} \cdot \mathbf{M}^{\frac{1}{2}} \right)^{T} \end{aligned} \tag{5.31}$$

Based on Equations 5.12 and 5.31, we can define

$$\mathbf{L} = \boldsymbol{\sigma}_{\mathbf{p}} \cdot \mathbf{V} \cdot \mathbf{M}^{\frac{1}{2}}$$

Correspondingly we can define the linear transformation by Equation 5.10 as

$$\mathbf{p} = \left(\boldsymbol{\sigma} \cdot \mathbf{V} \cdot \mathbf{M}^{\frac{1}{2}} \right) \cdot \mathbf{Z} + E(\mathbf{p}) \tag{5.32}$$

where $\mathbf{Z} = (Z_1, \cdots, Z_N)^T$ is a column vector of independent standard normal random variables as defined in Equations 5.7-5.9.

Based on the linear transformation, we can generate independently the samples of the standard normal random variables $Z_k \sim N(0,1)$, $k = 1, \cdots, N$, by the standard algorithms [Kurowicka and Cooke 2006]. The samples of **p** can then be generated by Equation 5.10 based on the samples of $\mathbf{Z}$.

Based on the linear transformation, we can also develop the analytical solution of the mean of the product of the correlated normal random variables. It will then be used

for calculating the analytical solution of the variance of the mimic model of the Fault-tree and Event-tree model.

5.4 Mean of Product of Correlated Normal Random Variables

Suppose that a set of normal random variables are defined as in Equations 5.4-5.6. We want to calculate the mean of the product $\prod_{i=1,N} p_i$, which is to be used for calculating the variance of the mimic model of the Fault-tree and Event-tree model. We are going to develop the analytical solution based on the linear transformation as defined in Equation 5.10.

We define the i^{th} row vector of the transformation matrix $\mathbf{L}$ as

$$\mathbf{L}_i = \left[\ell_{i,1}, \quad \cdots, \quad \ell_{i,k}, \quad \cdots, \quad \ell_{i,N}\right] \tag{5.33}$$

Therefore we have

$$\mathbf{L} = \begin{bmatrix} \mathbf{L}_1 \\ \vdots \\ \mathbf{L}_i \\ \vdots \\ \mathbf{L}_N \end{bmatrix}$$

Based on Equations 5.10 and 5.33, we have

$$p_i = \mathbf{L}_i \cdot \mathbf{Z} + E(p_i) = \sum_{k=1}^{N} \ell_{i,k} \cdot Z_k + E(p_i) \tag{5.34}$$

Consequently we have

$$\prod_{i=1,N} p_i = \prod_{i=1,N} \left(\sum_{k=1}^{N} \ell_{i,k} \cdot Z_k + E(p_i) \right) \tag{5.35}$$

Therefore the product $\prod_{i=1,N} p_i$ can be defined as a polynomial expansion as

$$\prod_{i=1,N} p_i = \sum_j T_j \tag{5.36}$$

and

$$T_j = \alpha_j Z_1^{\beta_{j,1}} \cdots Z_k^{\beta_{j,k}} \cdots Z_N^{\beta_{j,N}} \tag{5.37}$$

where α_j is the coefficient; $\beta_{j,k}$'s are the integer exponents.

In Equation 5.37, T_j is the product of *N* terms, one and only one of which is either $\ell_{i,k} Z_k$ or $E(p_i)$ that is associated with each p_i as defined in Equation 5.34. We define the index

$$s(j,i) = \begin{cases} k, \text{if } \ell_{i,k} Z_k \text{ is a term included in } T_j \\ 0, \text{if } E(p_i) \text{ is a term included in } T_j \end{cases} \tag{5.38}$$

Therefore we have $s(j,i) \in [0,N]$. We define the delta function

$$\delta(c,k) = \begin{cases} 1, \text{if } k = c \\ 0, \text{otherwise} \end{cases} \tag{5.39}$$

Based on Equations 5.35-5.39, we have

$$T_j = \prod_{i=1,N} \left(\delta(0, s(j,i)) \cdot E(p_i) + (1 - \delta(0, s(j,i))) \cdot \ell_{i,s(j,i)} \cdot Z_{s(j,i)} \right) \tag{5.40}$$

Applying Equation 5.40 with Equation 5.37, we have

$$\alpha_j = \prod_{i=1,N} \left(\delta(0, s(j,i)) \cdot E(p_i) + (1 - \delta(0, s(j,i))) \cdot \ell_{i,k(i,j)} \right) \tag{5.41}$$

and

$$\beta_{j,k} = \sum_{i=1}^{N} \delta(k, s(j,i)) \tag{5.42}$$

By Equation 5.39, we have

$$0 \le \beta_{j,k} \le N$$

With α_j and $\beta_{j,k}$ defined in Equations 5.41 and 5.42, we continue to derive the mean of $\prod_{i=1,N} p_i$. Based on Equation 5.36, we have

$$E\left(\prod_{i=1,N} p_i \right) = \sum_j E\left(\alpha_j Z_1^{\beta_{j,1}} \cdots Z_N^{\beta_{j,N}} \right)$$

Recalling that $Z_1, \cdots, Z_N$ are independent, we have

$$E\left(\prod_{i=1,I} p_i \right) = \sum_j \alpha_j E\left(Z_1^{\beta_{j,1}} \right) \cdots E\left(Z_N^{\beta_{j,N}} \right) \tag{5.43}$$

In Equation 5.43, $E\left(Z_k^{\beta_{j,k}} \right)$ represents the $\beta_{j,k}$-order moment of Z_k. Because Z_k is a standard normal random variable, this moment can be calculated as given in [Johnson and Balakrishnan 1994] as:

$$E\left(Z_k^{2i-1} \right) = 0, \; i = 1,2 \cdots \tag{5.44.a}$$

$$E\left(Z_k^{2i}\right) = (2i-1)(2i-3)\cdots 3\cdot 1,\ \ i = 1,2\cdots \tag{5.44.b}$$

The moment $E\left(Z_k^{\beta_{j,k}}\right)$ can also be calculated through an iterative process [Johnson and Balakrishnan 1994] as

$$E\left(Z_k^0\right) = 1 \tag{5.45.a}$$

$$E\left(Z_k^{2i}\right) / E\left(Z_k^{2(i-1)}\right) = (2i-1),\ i = 1,2\cdots \tag{5.45.b}$$

We have $E\left(Z_k^{2i-1}\right) = 0$ because the normal distribution is symmetric.

Some moments of a standard normal random variable are calculated and summarized in Table 5.2. Therefore Equation 5.43 offers a way to calculate the mean of the product of correlated normal random variables. This formula is to be used in calculating the variance of the mimic model of the Fault-tree and Event-tree model.

Table 5.2 Moments of a standard normal random variable

Order *i*	**Moment** $E\left(Z^i\right)$
0	1
1	0
2	1
3	0
4	3
5	0
6	15
7	0
8	105

5.5 Mean of Product of two Correlated Normal Random Variables

Suppose that p_i and p_j are two correlated normal random variables. The mean of the product of p_i and p_j can be calculated directly rather than through the linear transformation as developed above.

The correlation between p_i and p_j is defined as

$$r(p_i, p_j) = \frac{\mathrm{cov}(p_i, p_j)}{\sqrt{\mathrm{var}(p_i)\mathrm{var}(p_j)}} \tag{5.46}$$

where the covariance is defined as

$$\mathrm{cov}(p_i, p_j) = E(p_i \cdot p_j) - E(p_i \cdot)E(p_j) \tag{5.47}$$

Applying Equation 5.47 with Equation 5.46, we have

$$E(p_i p_j) = E(p_i)E(p_j) + r(p_i, p_j)\sqrt{\mathrm{var}(p_i)\mathrm{var}(p_j)} \tag{5.48}$$

Based on Equation 5.48, the mean of the product of two correlated normal random variables can be calculated based on the means, the variance and correlation. This analytical solution will be used to calculate the analytical solution of the variance of the mimic model.

5.6 Analytical Solution of the Variance of the Mimic Model

5.6.1 Overview

When rare event approximation is applied [Bedford and Cooke 2001], the mimic model of a Fault-tree and Event-tree model can be approximated as in Equations 5.1-5.3. Based on Equation 5.3, we have

$$E(r) = \sum_{\ell=1}^{N(S)} E(r_\ell) \tag{5.49}$$

and

$$E(r^2) = E\left(\sum_{\ell(1)=1}^{N(S)} \sum_{\ell(2)=1}^{N(S)} (r_{\ell(1)} \cdot r_{\ell(2)}) \right) = \sum_{\ell(1)=1}^{N(S)} \sum_{\ell(2)=1}^{N(S)} E(r_{\ell(1)} \cdot r_{\ell(2)}) \tag{5.50}$$

where $N(S)$ is the number of the risk scenarios.

We know that the variance of the risk *r* can be calculated by

$$\text{var}(r) = E(r^2) - E(r) \cdot E(r) \tag{5.51}$$

Therefore to calculate $\text{var}(r)$, we need to calculate $E(r_\ell)$ and $E(r_{\ell(1)} \cdot r_{\ell(2)})$. We assume that the occurrence frequency f_ℓ and the consequence c_ℓ are independent. Consequently from Equation 5.2 we have

$$E(r_\ell) = E(f_\ell) \cdot E(c_\ell) \tag{5.52}$$

$$E(r_{\ell(1)} \cdot r_{\ell(2)}) = E(f_{\ell(1)} \cdot f_{\ell(2)}) \cdot E(c_{\ell(1)} \cdot c_{\ell(2)}) \tag{5.53}$$

From Equations 5.52 and 5.53, to calculate $\text{var}(r)$, we need to calculate $E(f_\ell)$, $E(c_\ell)$, $E(f_{\ell(1)} \cdot f_{\ell(2)})$ and $E(c_{\ell(1)} \cdot c_{\ell(2)})$. The methods are to be developed later on.

5.6.2 Calculation of $E(c_\ell)$ and $E(c_{\ell(1)} \cdot c_{\ell(2)})$

When a simplified model of two injury levels is applied, the individual casualty probability is calculated for each rule set as developed in Chapter 4. Within the

casualties, a ratio of the fatalities to the major injuries can be set. The consequence can then be calculated as the equivalent fatalities that counts both the fatalities and the scaled-down major injuries. We denote by ϕ the proportion of the fatalities in he casualties. We suppose that W major injuries are counted as one equivalent fatality. The equivalent fatalities of S_ℓ can then be defined as the mean as

$$c_\ell = N_\ell \times \pi_i(4) \times (\phi + (1-\phi)/W) \tag{5.54}$$

where N_ℓ stands for the mean of the number of people exposed to the risk scenario S_ℓ; $\pi_i(4)$ stands for the individual casualty probability of the rule set R_i that is associated with S_ℓ.

Because a rule set can be associated with multiple consequences, the subscript of the rule set is not necessarily equal to the subscript of the consequence in Equation 5.54.

As discussed previously, in this research, we don't study the uncertainties in N_ℓ, ϕ, and W. The uncertainty in the consequence c_ℓ is then driven by the uncertainty in the associated rule set only. Consequently based on Equation 5.54, we have

$$E(c_\ell) = N_\ell \times (\phi + (1-\phi)/W) \times E(\pi_i(4)) \tag{5.55}$$

In Equation 5.55, the individual casualty probability $\pi_i(4)$ has been defined on the injury atoms as in Equation 4.67 as

$$\pi_i(4) = 1 - \prod_{j=1}^{D} \left(1 - \frac{z_{i,j}(1,4)}{\gamma_j} \cdot A_j\right) \tag{5.56}$$

where A_i stands for the i^{th} injury atom; D stands for the number of the injury atoms.

The injury atoms A_j have been assumed to be independent. Consequently by Equation 5.56, we have

$$E(\pi_i(4)) = 1 - \prod_{j=1}^{D} \left(1 - \frac{z_{i,j}(1,4)}{\gamma_j} \cdot E(A_j) \right) \tag{5.57}$$

We have assumed that the uncertainty in the injury atom A_j follows a Beta distribution as defined in Equation 4.71. The mean of the injury atom $E(A_j)$ can then be calculated by

$$E(A_j) = \frac{\alpha_j}{\alpha_j + \beta_j} \tag{5.58}$$

The definition parameters α_j and β_j can be derived for A_j through elicitation by the method developed in Chapter 4. Applying Equations 5.58 and 5.57 with Equation 5.55, we can calculate the mean of the consequence c_ℓ. We continue to derive the mean of the product of two consequences.

Suppose there are two consequences $c_{\ell(1)}$ and $c_{\ell(2)}$ that are associated with rule set $R_{i(1)}$ and $R_{i(2)}$ respectively. By Equation 5.54, we have

$$E(c_{\ell(1)} \cdot c_{\ell(2)}) = N_{\ell(1)}(P) \cdot N_{\ell(2)}(P) \cdot (\phi + (1-\phi)/W)^2 \cdot E(\pi_{i(1)}(4) \cdot \pi_{i(2)}(4)) \tag{5.59}$$

We need to calculate the mean $E(\pi_{i(1)}(4) \cdot \pi_{i(2)}(4))$. By Equation 5.56, we have

$$\pi_{i(1)}(4)\cdot\pi_{i(2)}(4)=\left(1-\prod_{j=1}^{D}\left(1-\frac{z_{i(1),j}(1,4)}{\gamma_j}\cdot A_j\right)\right)\cdot\left(1-\prod_{j=1}^{D}\left(1-\frac{z_{i(2),j}(1,4)}{\gamma_j}\cdot A_j\right)\right)$$

$$=1-\prod_{j=1}^{D}\left(1-\frac{z_{i(1),j}(1,4)}{\gamma_j}\cdot A_j\right)-\prod_{j=1}^{D}\left(1-\frac{z_{i(2),j}(1,4)}{\gamma_j}\cdot A_j\right)$$
$$+\prod_{j=1}^{D}\left(1-\frac{z_{i(1),j}(1,4)}{\gamma_j}\cdot A_j\right)\prod_{j=1}^{D}\left(1-\frac{z_{i(2),j}(1,4)}{\gamma_j}\cdot A_j\right)$$

(5.60)

Applying Equation 5.56 with Equation 5.60, we have

$$\pi_{i(1)}(4)\cdot\pi_{i(2)}(4)=\pi_{i(1)}(4)+\pi_{i(2)}(4)-1$$
$$+\prod_{j=1}^{D}\left(1-\frac{z_{i(1),j}(1,4)+z_{i(2),j}(1,4)}{\gamma_j}\cdot A_j+\frac{z_{i(1),j}(1,4)\cdot z_{i(2),j}(1,4)}{\gamma_j^2}\cdot A_j^2\right)$$

(5.61)

Each injury atom is defined for one hazard type as given in Chapter 4. The injury atoms have also been assumed to be independent, i.e. a known transition probability of one injury atom does not tell any information on the transition probability of another injury atom. Therefore, we have

$$E\left(\pi_{i(1)}(4)\cdot\pi_{i(2)}(4)\right)$$

$$=E\left(\pi_{i(1)}(4)\right)+E\left(\pi_{i(2)}(4)\right)-1$$
$$+\prod_{j=1}^{D}\left(1-\frac{z_{i(1),j}(1,4)+z_{i(2),j}(1,4)}{\gamma_j}\cdot E\left(A_j\right)+\frac{z_{i(1),j}(1,4)\cdot z_{i(2),j}(1,4)}{\gamma_j^2}\cdot E\left(A_j^2\right)\right)$$

(5.62)

In Equation 5.62, the mean of the injury atom $E(A_j)$ can then be calculated by Equation 4.72 as discussed above. Because the injury atom A_j is assumed to follow

a Beta distribution as defined in Equation 4.71, we can calculate the variance of A_j by

$$\mathrm{var}(A_j) = \frac{\alpha_j \beta_j}{(\alpha_j + \beta_j)^2 (\alpha_j + \beta_j + 1)} \tag{5.63}$$

where α_j and β_j are the definition parameters of A_j.

Based on Equations 5.58 and 5.63, we have

$$\begin{aligned} E(A_j^2) &= \mathrm{var}(A_j) + E(A_j) \cdot E(A_j) \\ &= \frac{\alpha_j \beta_j}{(\alpha_j + \beta_j)^2 (\alpha_j + \beta_j + 1)} + \left(\frac{\alpha_j}{\alpha_j + \beta_j} \right)^2 \end{aligned} \tag{5.64}$$

By applying Equations 5.62, 5.58 and 5.64 with Equation 5.59, we can calculate the mean of the product of two consequence, i.e. $E(c_{\ell(1)} \cdot c_{\ell(2)})$. The means $E(c_\ell)$ and $E(c_{\ell(1)} \cdot c_{\ell(2)})$ can then be used to calculate the mean and the variance of the output of the Fault-tree and Event-tree model by Equations 5.49-5.53.

We continue to develop the calculation of $E(f_\ell)$ and $E(f_\ell \cdot f_k)$ for calculating the analytical solution of the variance of the output of Fault-tree and Event-tree models by Equations 5.52 and 5.53.

5.6.3 Calculation of $E(f_\ell)$ and $E(f_{\ell(1)} \cdot f_{\ell(2)})$

As defined in Equation 5.1, the frequency of a risk scenario is the product of the associated basic events and the probabilities of the associated outcomes of the escalation events. As stated in Chapter 4, in this research we study the uncertainty of the means of the basic events and treat the probabilities of the outcomes of the escalation event as constant. Therefore based on Equation 5.1, we have

$$E(f_\ell) = E\left(\prod_{p_i \in S_\ell} p_i\right) \cdot \prod_{e(\ell,i) \in S_\ell} \Pr(e(\ell,i)) \tag{5.65}$$

where p_i stands for a basic event; $\Pr(e(\ell,i))$ stands for the probability of the outcome $e(\ell,i)$ of the escalation event E_i associated with the risk scenario S_ℓ.

For the product of the frequencies of two risk scenarios, we have

$$E(f_{\ell(1)} \cdot f_{\ell(2)}) = E\left(\prod_{p_i \in S_{\ell(1)}} p_i \prod_{p_j \in S_{\ell(2)}} p_j\right) \cdot \prod_{e(\ell(1),i) \in S_{\ell(1)}} \Pr(e(\ell(1),i)) \cdot \prod_{e(\ell(2),j) \in S_{\ell(2)}} \Pr(e(\ell(2),j)) \tag{5.66}$$

We have assumed that the basic events p_i's follow a joint normal distribution. Therefore $E\left(\prod_{p_i \in S_\ell} p_i\right)$ and $E\left(\prod_{p_i \in S_{\ell(1)}} p_i \prod_{p_j \in S_{\ell(2)}} p_j\right)$ can be calculated by Equation 5.43. The means $E(f_\ell)$ and $E(f_{\ell_1} f_{\ell_2})$ can be calculated consequently.

Suppose that in Equation 5.66, all the basic events $p_i \in S_{\ell(1)}$ are independent and all the basic events $p_j \in S_{\ell(2)}$ are independent. Consequently one basic event $p_m \in S_{\ell(1)}$ can be correlated with at most one basic event $p_n \in S_{\ell(2)}$. Accordingly $E\left(\prod_{p_i \in S_{\ell(1)}} p_i \prod_{p_j \in S_{\ell(2)}} p_j\right)$ can be expressed of the form

$$E\left(\prod_{p_i \in S_{\ell(1)}} p_i \prod_{p_j \in S_{\ell(2)}} p_j\right) = \prod E(p_k) \prod_{\substack{p_m \in S_{\ell(1)} \\ p_n \in S_{\ell(2)}}} E(p_m p_n) \tag{5.67}$$

where p_m and p_n stand for a pair of dependent basic events one from each risk scenario; p_k stands for an independent basic event that can be from either risk scenario.

Consequently, the mean $E(p_m p_n)$ can be calculated by Equation 5.48 that represents a simpler way than that represented by Equation 5.43. As a result, the mean $E\left(\prod_{p_i \in S_{\ell(1)}} p_i \prod_{p_j \in S_{\ell(2)}} p_j\right)$ can also be calculated in a simpler way.

As a summary, after $E(f_\ell)$, $E(c_\ell)$, $E(f_{\ell(1)} \cdot f_{\ell(2)})$ and $E(c_{\ell(1)} \cdot c_{\ell(2)})$ are calculated, the variance $\text{var}(r)$ can be calculated by Equations 5.49-5.53.

5.7 Software Design

5.7.1 Overview

The whole mission is carried out with three Excel workbooks designated for: (1) building the uncertainty model in the rule sets; (2) building the uncertainty model for the basic events; and (3) calculating the analytical variance and conducting simulations. Based on the methods developed in the previous sections, the software design for calculating the variance of the mimic models is divided into three major procedures: (1) the console procedure; (2) the procedure for calculating the mean of the consequence and the mean of the product of two consequences; (3) the procedure for calculating the mean of the frequency and the mean of the product of two frequencies. The three procedures are implemented by programming with Excel VBA associated with the three Excel workbooks. We then discuss the methods for validating the software.

5.7.2 Console Procedure

The console procedure is designed based on Equations 5.50-5.51 to calculate the variance of the mimic model of the Fault-tree and Event-tree models. It is

implemented by the function *outputVar()*. As shown in Fig. 5.3, the function *outputVar()* is implemented with two loops. The outer loop is designed to go through all the risk scenarios included in the mimic model as demonstrated in Fig. 5.2. The loop variable *i* changes from 1 to *N(S)* that stands for the risk scenario number. For each risk scenario, the mean of the consequence EF_i is calculated by Equations 5.55 and 5.24; the mean of the frequency EFQ_i is calculated by Equations 5.65 and 5.43. The product of EFQ_i and EF_i represents the mean of the equivalent fatalities associated with the current risk scenario as defined in Equation 5.52, which is added up to the variable *ESCS* standing for the expectation of the sum of the risk scenario. The value of *ESCS* after the loop conveys the mean $E(r)$ as defined in Equation 5.49.

For each risk scenario, the mean of the square of the consequence EF_{ii} is calculated by Equations 5.59 and 5.62; the mean of the square of the frequency EFQ_{ii} is calculated by Equations 5.66 and 5.43. The product of EFQ_{ii} and EF_{ii} represents the mean of the square of the equivalent fatalities associated with the current risk scenario, which is added up to the variable *ESCSS* standing for the expectation of the sum of the risk scenario square.

As shown in Fig. 5.3, the inner loop is designed to calculate the mean of the product of two different risk scenarios. The loop variable *j* changes from *i* to *CN*. Within the loop, the mean of the product of the i^{th} risk scenario frequency and the j^{th} risk scenario frequency, denoted as $EFQ_{ij,}$ is calculated based on Equations 5.66 an 5.43; the mean of the product of the i^{th} risk scenario consequence and the j^{th} risk scenario consequence, denoted as EF_{ij} is calculated based on Equations 5.59 and 5.62. The product of EFQ_{ij} and EF_{ji} represents the mean of the product of the i^{th} risk scenario risk and the j^{th} risk scenario risk as defined in Equation 5.53, which is doubled and then added up to *ESCSS*. The value of *ESCSS* after the both loops conveys the mean $E(r^2)$ as defined in Equation 5.50.

At the end of the procedure, the variance of the mimic model output can be calculated as $ESCSS - ESCS \times ESCS$ by the Equation 5.51.

As shown in Fig. 5.3, the function *outputVar()* calls two functions *CutSetMean*(*i*,*j*) and *consequencesMean*(*i, j*). The function *CutSetMean*(*i*,*j*) calculates the mean of the product of two risk scenario frequencies. By setting $j = 0$, the mean of the i^{th} risk scenario frequency can be calculated by the same function. The function *consequencesMean*(*i, j*) calculates the mean of the of the product of two risk scenario consequences. By setting $j = 0$, the mean of the i^{th} risk scenario consequence can be calculated by the same function. The flowcharts of the two functions are to be defined separately as follows.

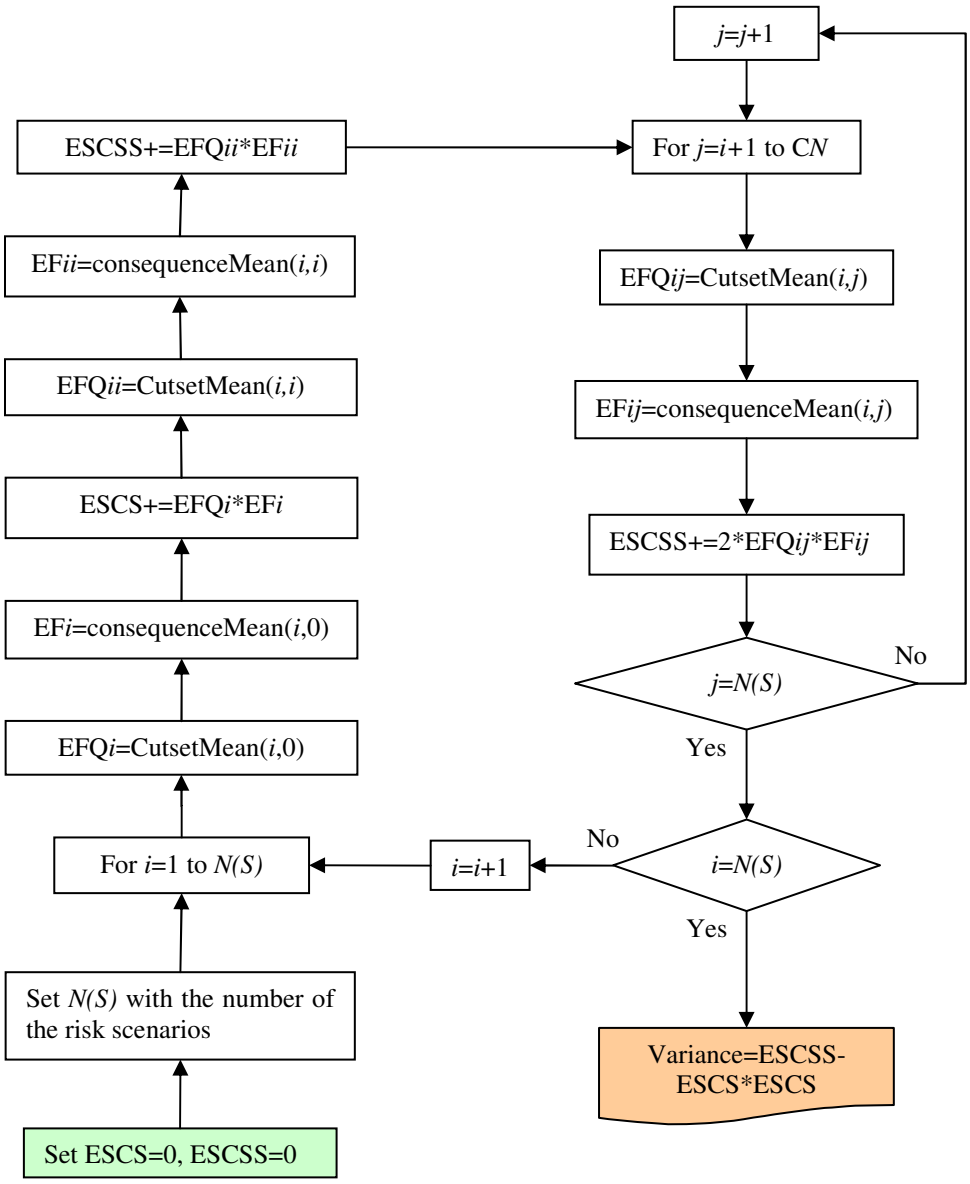


Fig. 5.3 Flowchart of Function *outputVar()* for calculating the variance of the mimic model of the Fault-tree and Event-tree models

5.7.3 Procedure for Calculating the Mean of the Consequence and the Mean of the Product of two Consequences

As shown in Fig. 5.3, the console procedure calls the function *consequenceMean(i, j)* for calculating the mean of the consequence and the mean of the product of two consequences. The flowchart of the procedure is demonstrated in Fig. 5.4. When $j = 0$ the procedure *consequenceMean(i, j)* calculates the mean of the consequence associated with the i^{th} risk scenario by Equation 5.55. When $j \geq i > 0$, the procedure *consequenceMean(i, j)* calculates the mean of the product of the two consequences associated with the i^{th} risk scenario and the j^{th} risk scenario by Equation 5.59. The procedure *consequenceMean(i, j)* requires two the inputs: (1) the means of the rule sets, defined *RSMeans(1 To RSN);* (2) the means of the products of a pair of rule sets *RSCPMeans(1 To RSN, 1 To RSN)*, where *RSN* stands for the number of the rule sets; *RSMeans(1 To RSN)* is defined as a one-dimensional data vector; *RSCPMeans(1 To RSN, 1 To RSN)* is defined as a two-dimensional data matrix. The two values *RSMeans* and *RSCPMeans* are calculated in the procedure *consequenceMean(i,j)* as illustrated in Fig. 5.5.

As illustrated in Fig. 5.5, the procedure *RuleSetsMeans()* start with reading in three data sets: *RSIA(1 to RSN, 1 to IAN), IADef(1 to IAN, 1 to 2)* and *IAMeans(1 to IAN, 1 to 2)*, where *IAN* stands for the number of the injury atoms. The data set *RSIA* is a two-dimensional matrix comprising the definition information of the rule sets on the injury atoms. Analogous to Table 4.3, the matrix *RSIA* composes of the weight coefficients $z_{k,j}$ and each row of *RSIA* corresponds to one rule set defined by Equation 4.9. The data set *IADef* is a two-dimensional matrix of the definition parameters of the injury atoms. Analogous to Table 4.2, each row of *RSIA* corresponds to one injury atom; and the two columns contain the parameters α and β of a Beta distribution respectively. The data set *IAMeans* is a two-dimensional matrix. With each row corresponding to one injury atom, the two columns of *IAMeans* contain the mean of the injury atom and the mean of the square of the injury atom that can be calculated by Equations 5.58 and 5.64 respectively.

The function *RuleSetsMeans()* then goes into the first loop to calculate the means of the rule sets by Equation 5.58. The results are stored in the one-dimensional matrix *RSMeans(1 to RSN),* with the cell *RSMeans(RSi)* contains the mean of the *RSi*th rule set. The function then enters into an embedded loop to calculate the means of the product of two rule sets by Equation 5.64. The results are stored in the two-dimensional matrix *RSCPMeans(1 to RSN, 1 to RSN),* with the cell *RSCPMeans(RSi, RSj)* contains the mean of the product of the *RSi*th and the *RSj*th rule set. The data sets *RuleSetsMeans* and *RSCPMeans* are then called in the function *consequenceMean(i, j)* to calculate the means of the consequence and the product of two consequence.

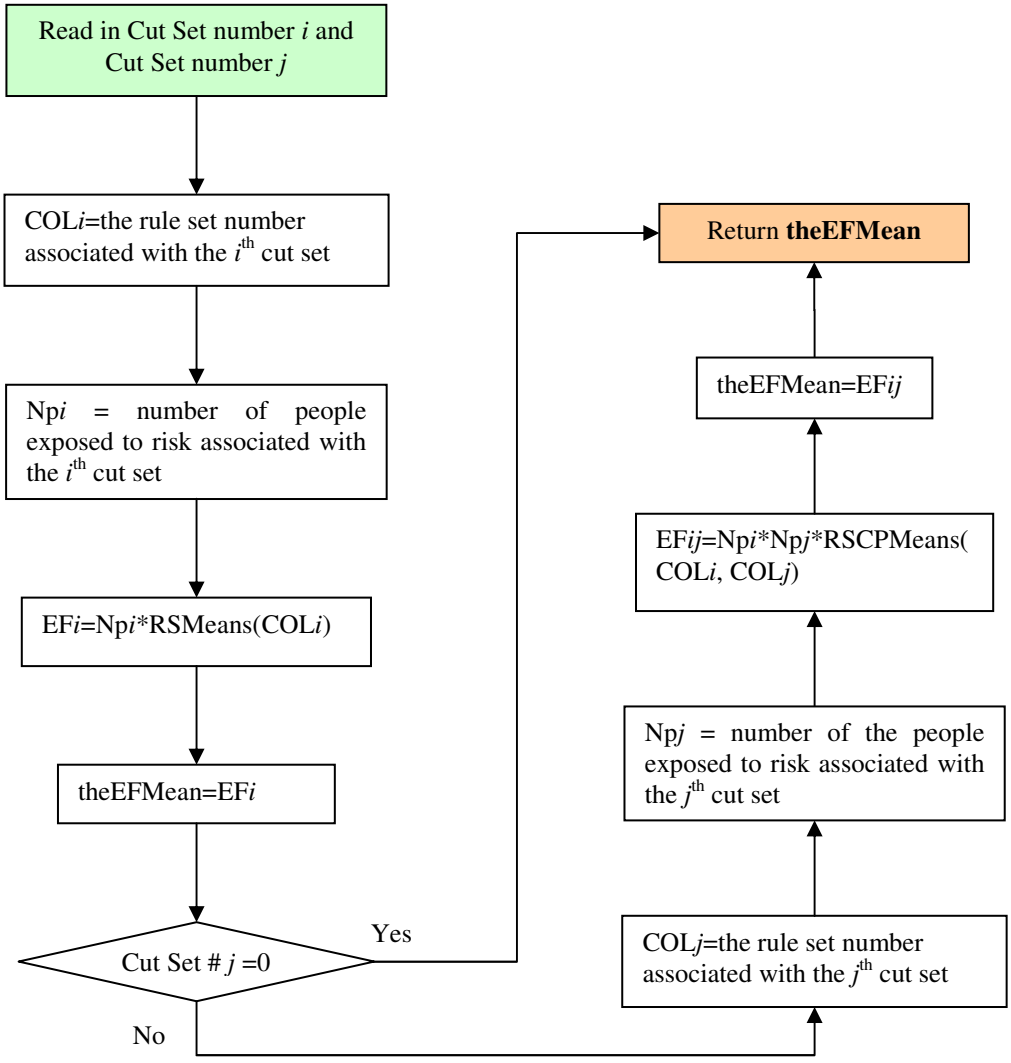


Fig. 5.4 Flowchart of Function *consequenceMean(i,j)* for calculating the mean of the i^{th} risk scenario consequence when $j = 0$ or the mean of the product of the i^{th} risk scenario consequence and the j^{th} risk scenario consequence when $j \geq i > 0$

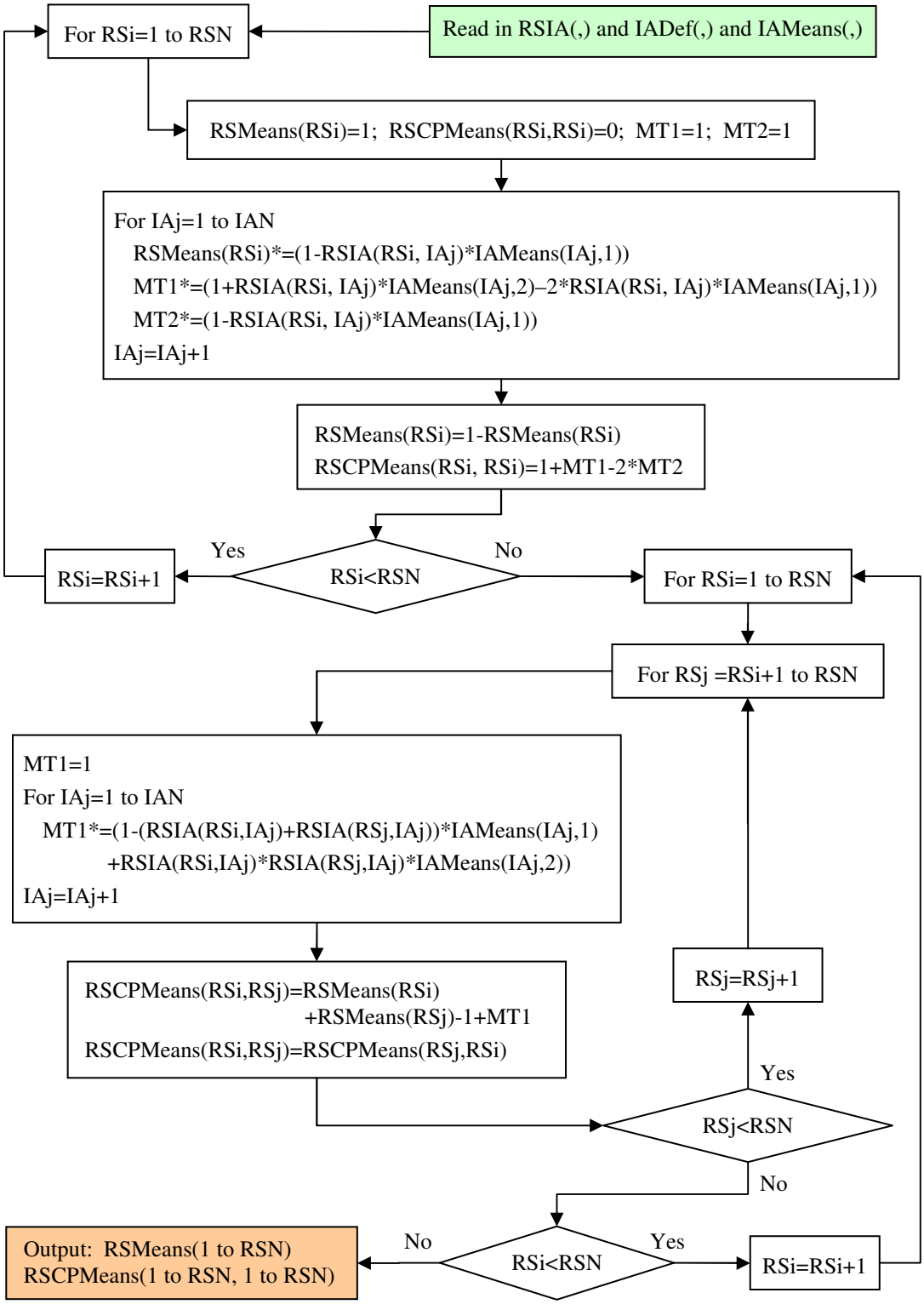


Fig. 5.5 Flowchart of the function *RuleSetsMeans()* for calculating the means of the rule set and mean of the product of two rule sets through injury atoms

5.7.4 Procedure for Calculating the Mean of the Frequency and the mean of the Product of two Frequencies

The procedure for calculating the mean of the frequency and the mean of the product of two frequencies is implemented in the function *CutsetMean(i, j).* It is called in the function console procedure *outputVar()* as illustrated in Fig. 5.3.

In the case studies [Harrison, Griffin et al. 2008] of this research, all the basic events composing one risk scenario are independent. Accordingly the procedure *CutsetMean(i, j)* is designed based on Equation 5.67. As illustrated in Fig. 5.6, the procedure *CutsetMean(i, j)* has two input parameters *i* and *j* that are the number of the risk scenarios in the mimic model as shown in Fig. 5.2. The procedure starts with reading in the codes of the basic events composing the i^{th} risk scenario and, when $j \neq 0$, those codes associated with the j^{th} risk scenario. These event codes are stored in the vector *ECodes(1 to EN)*, where *EN* stands for the number of the read-in events. Corresponding to the events defined in *ECodes(1 to EN),* the correlation matrix, the standard deviation and the means are then read into *corr(1 to EN, 1 to EN)*, *std(1 to EN)* and *EMeans(1 to EN)* respectively.

The procedure *CutsetMean(i, j)* with initializes *theMean=1* and *Done(1 to EN)=0*. It then enters into the loop for *Ei=1 to EN*. When *Done(Ei)=0,* the mean *EMeans(Ei)* is recorded. The procedure then goes into the embedded loop for *Ej=Ei+1 to EN*. The correlation *corr(Ei,Ej)* is checked. When *corr(Ei,Ej)≠0*, the event *ECodes(Ej)* is the only one correlated with the event *ECodes(Ei).* The mean of the product of *ECodes(Ej)* and *ECodes(Ei)*, i.e. $E(p_i p_j)$, is calculated by Equation 5.48. The procedure *CutsetMean(i, j)* then marks the event *ECodes(Ej)* as having been dealt with by setting *Done(Ej)=1*. After that, the procedure jumps out of the embedded loop. Depending if there exists a correlated event *ECodes(Ej)*, the output mean *theMean* is multiplied by *EMeans(Ei) or* $E(p_i p_j)$.

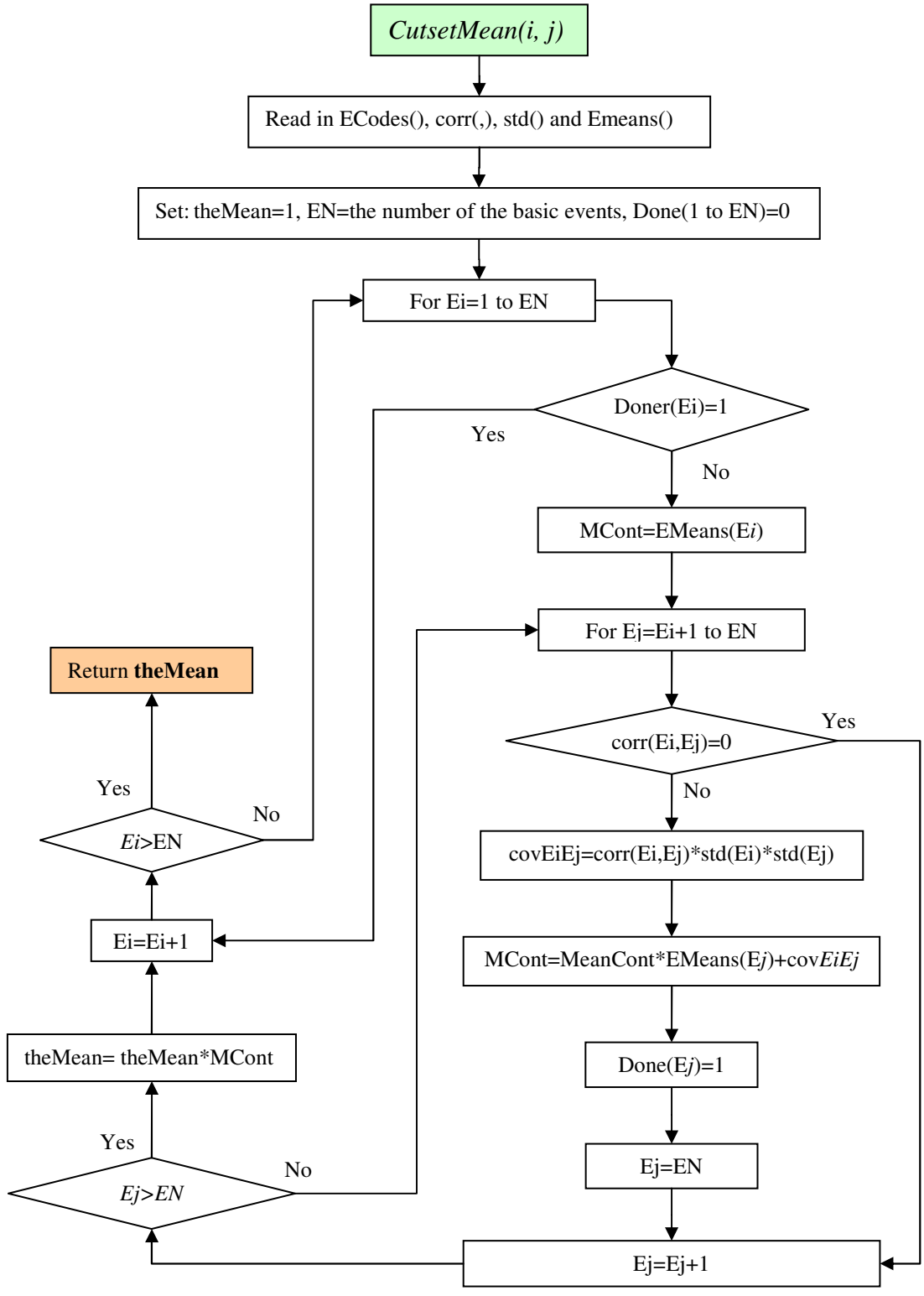


Fig. 5.6 Flowchart of the procedure *CutsetMean(i, j)* for calculating the mean of the frequency and the mean of the product of two frequencies

5.7.5 Software Implementation

5.7.5.1 Overview

Based on the mimic model as illustrated in Fig. 5.2, the software for calculating the variance can be developed with MS Excel. According to Fig. 5.1, the software can be partitioned into three parts defined as: (1) Part 1 for building the uncertainty model in the rule sets corresponding to Block 6; (2) Part 2 for building the uncertainty model of the precursors corresponding to Block 4; and (3) Part 3 for conducting uncertainty analysis corresponding to the rest blocks. Parts 1 to 3 are implemented in three Excel workbooks named as *rulesetsUM.xls, precursorsUM.xls* and *UAnalysis.xls* respectively. When needed, the data stored in the three workbooks can be accessed from any other ones by programming with Excel VBA [Jelen and Syrstad 2004].

5.7.5.2 Workbook *rulesetsUM*

Workbook *rulesetsUM.xls* composes of 4 parts including:

- the definition of the injury atoms;
- the definition of the rule sets on the injury atoms;
- the means of the rule sets and the products of two rule sets;
- the samples of the rule sets

As demonstrated in Fig. 5.7, the injury atoms are defined with the code, the definition parameters Alpha and Beta, the mean and the standard deviation. The codes are uniquely defined and used to access to the injury atoms. The definition of the rule sets on the injury atoms is demonstrated in Fig. 5.8. Based on these two definitions, the means of the rule sets and the means of the products of two rule sets can be calculated by programming with VBA. It is implemented by the function *RuleSetsMeans()* as illustrated in Fig. 5.5. The outcome means of the rule sets and the means of the products of two rule sets can be stored in a worksheet in Workbook *'rulesetsUM.xls",* which can then be accessed by the function *consequenceMeans(i,j)* for calculating the means of the consequences and the means of the products of the consequences, as illustrated in Fig. 5.4.

The samples of the rule sets can also be generated by programming with Excel VBA in Workbook *rulesetsUM.xls*. The samples are stored in a worksheet of Workbook *rulesetsUM.xls* and to be accessed from Workbook *UAnalysis.xls* for conducting simulations.

Injury Atom Code	Alpha	Beta	mean	std
DERTNL	7.9486	1846.6	0.004286003	0.001516552
CARONSTNL	5.025	15.676	0.242741897	0.092035306
STROnTNOS	4.9327	14.798	0.250001267	0.095103037
SecCollTNL	5.3845	19.743	0.214287136	0.080275137
FireTNL	20.309	2.2566	0.899998227	0.061799566
ToxicTNL	20.309	2.2566	0.899998227	0.061799566
Onbridge	6.2706	37.624	0.142855841	0.052225052

Fig. 5.7 Excel worksheet layout of the definition of the injury atoms

Rule Sets	DERTNL	CARONSTNL	STROnTNOS	SecCollTNL	FireTNL	ToxicTNL	OnBridge
T12-BG-1	0	0	0	0	0	0	1
T12-OF-01	0.333333333	0	0	0	0	0	0
T12-OF-02	0.333333333	0	0	0	0	0	0
T12-OF-03	0.333333333	0	0	0.045684307	0	0	0
T12-OF-04	0.333333333	0	0	0.045684307	0.031746032	0	0
T12-OF-05	0.333333333	0	1	0.045684307	0	0	0
T12-OF-06	0.333333333	0	1	0.045684307	0.031746032	0	0
T12-OF-07	0.333333333	0.54096812	0	0	0	0	0
T12-OF-08	0.333333333	0.54096812	0	0	0.031746032	0	0
T12-OF-09	0.333333333	0.54096812	0	0.133333333	0	0	0
T12-OF-10	0.333333333	0.54096812	0	0.133333333	0.031746032	0	0
T12-OF-11	0.333333333	0.54096812	0.657142857	0.133333333	0	0	0
T12-OF-12	0.333333333	0.54096812	0.657142857	0.133333333	0.047619048	0	0

Fig. 5.8 Excel worksheet layout of the definition of the rule sets on the injury atoms

5.7.5.3 Workbook *precursorsUM*

Workbook *precursorsUM.xls* contains the covariance matrix and the means of the precursors. As shown in Fig. 5.9 the codes of the precursors are put on the top row of the worksheet. These codes are uniquely defined and are used to access the variance. The map from the precursor codes to the associated column numbers can be built by programming. Suppose that there are two precursors PC1 and PC2 and the associated

column numbers are *NPC1* and *NPC2* respectively. The covariance between PC1 and PC2 can be accessed by either Cells(*NPC1+1*, *NPC2*) or Cells(*NPC2+1*, *NPC1*) for the symmetric covariance matrix [Jelen and Syrstad 2004].

The samples of the precursors can be generated by the methods developed in Section 5.3. These samples are stored in a separate worksheet in Workbook *precursorsUM.xls*, which are accessed during simulations.

POSL----	PI	PSNT----	PI	PSPD----	PI	UTRN----	U	BTRE----	U	RLNS----	U	RQAK----	U	RSLP----	PI	WFLD----	U	WSNO----	U	WWIN----	U
3E-21		1.56E-21		1.59E-20		1.59E-22		0		0		0		0		0		0		0	
1.56E-21		1.53E-21		9.36E-21		9.36E-23		0		0		0		0		0		0		0	
1.59E-20		9.36E-21		1.59E-19		9.54E-22		0		0		0		0		0		0		0	
1.59E-22		9.36E-23		9.54E-22		1.59E-23		0		0		0		0		0		0		0	
0		0		0		0		6.11E-20		1.68E-19		0		2.65E-20		7.08E-24		1.38E-20		8.42E-23	
0		0		0		0		1.68E-19		5.5E-19		0		7.94E-20		2.13E-23		4.14E-20		2.53E-22	
0		0		0		0		0		0		5.19E-27		0		0		0		0	
0		0		0		0		2.65E-20		7.94E-20		0		1.36E-20		3.34E-24		6.5E-21		3.97E-23	
0		0		0		0		7.08E-24		2.13E-23		0		3.34E-24		1.32E-27		1.61E-24		1.06E-26	
0		0		0		0		1.38E-20		4.14E-20		0		6.5E-21		1.61E-24		4.08E-21		2.07E-23	

Fig. 5.9 Excel worksheet layout of the covariance matrix of the precursors

5.7.5.4 Workbook *UAnalysis*

Workbook *UAnalysis.xls* contains the definitions of the mimic models as demonstrated in Fig. 5.2. The three functions *outputVar(), consequenceMean(i, j)* and *CutSetMean*(*i*,*j*) are implemented by programming with Excel VBA in this workbook for calculating the mean and variance of the mimic model. These functions access Workbook *precursorsUM.xls* and Workbook *rulesetsUM.xls* for the uncertainty models of the precursors and the rule sets respectively.

The simulations process is also implemented on Workbook *UAnalysis.xls*. The samples of the precursors and the rule sets are fetched from Workbook *precursorsUM.xls* and Workbook *rulesetsUM.xls* respectively.

5.8 Methods for Validating the Software

At the first stage, the mimic models need to be validated. We need to prove that the mimic model is set up correctly and therefore can be used for uncertainty analysis. This can be done by comparing the result from the mimic model with that from the original model. Given the same values of the input parameters, we run the original Fault-tree and Event-tree model and the mimic model separately. The outcome results are compared to see if the difference is less than the required precision.

At the second step, we need to test the methods and the software codes for calculating the variance of the mimic model. This can be done through a small mimic model of a few risk scenarios. For such a small mimic model, the analytical solutions of the mean and the variance can be calculated manually. The results can then be used to check the outcome results from running the mimic model. Although the mimic model composes of only a few risk scenarios, all the methods and the software codes have to be run in the same way as for a large real mimic model to calculate the mean and the variance. Therefore a small mimic model is efficient to test the methods and the software codes for calculating the variance.

At the third step, we need to test that the methods and the software codes can work robustly with large real mimic models. We know that unbiased estimations of the mean and the variance of the mimic model can be obtained from the Monte-Carlo simulations [Helton and Davis 2000; Kurowicka and Cooke 2006]. We can then compare the analytical solutions with the simulation results. Consistent results demonstrate that the methods and the software for analytical solutions are robust for large real mimic models.

5.9 Conclusions

In this chapter, a set of methods are developed for conducting uncertainty analysis of Fault-tree and Event-tree models built on the commercial software such as Isograph FT+. It is generally difficult to access and manipulate the data of such Fault-tree and Event-tree models [Isograph 2004; Isograph 2005]. It causes the problem in

conducting simulations for uncertainty assessment. Because the computer models of the Fault-tree and Event-tree models are the "black-box" to the user, it is impossible to conduct any kind of analytical analysis of the uncertainty in the output. To solve these problems, we develop a mimic MS Excel model for the original Fault-tree and Event-tree model. The first benefit is that the MS Excel models are completely transparent. The simulations on the mimic models are therefore easy to implement. We can also calculate the analytical variance of the mimic model. The second benefit of developing the mimic models is associated with the popularity of MS Excel especially in business related analysis. An Excel mimic model can be used easily by more analysts for various purposes.

Based on the mimic models, we use simulations to build up the empirical distribution of the risk. We also develop the method for calculating the analytical solution of the variance of the risk. These two methods are implemented independently and therefore can be used for cross check to each other. In these two methods, we need to cope with the dependent basic events which uncertainty is assumed to follow a joint normal distribution. As usually, the dependent basic events are expressed in terms of a set of independent standard normal random variables by linear transformation. In risk analysis context, however, there are two issues that affect the implementation of the linear transformation. First, the covariance matrix can be the positive semi-definite, to which the *Cholesky* decomposition is not suitable. To solve this issue, we select the linear transformation through eigen-decomposition. Second, the basic events may have very small standard deviation [Harrison, Griffin et al. 2008], which makes the calculation of the eigenvalues and the eigenvectors more expensive and exposed to larger errors. To solve this issue, we select the linear transformation through the correlation matrix.

We then design the software to implement the above methods. The software is implemented on MS Excel workbooks and therefore can be conveniently installed and run in applications. The software is validated at three stages including: (1) validate the mimic model; (2) validate correctness of the methods and the software; (3) validate the robustness of the methods and the software.

For future research work, we can compare the Eigen-decomposition of the correlation matrix to the factors elicited for building the correlation matrix. We can also continue to investigate the method for calculating the distribution of the sum of products of lognormal random variables, which can be used in the future.

Chapter 6

Case Studies with RSSB-SRM HET10 and HET12

6.1 Overview

In this chapter, we will conduct two case studies to assess the uncertainties in the output of a Fault-tree and Event-tree model. Through these case studies, we will test the practical performance of the procedure and the methods for uncertainty analysis that are developed in the previous chapters. For each case study, at the first step, we will build the covariance matrix of the basic events of the fault tree by the procedure developed in Chapter 3. We then build the uncertainty model in the rule sets of the event tree by the method developed in Chapter 4. With the uncertainty models in the basic events and the rule sets, we then assess the uncertainties in the output by the methods developed in Chapter 5.

The case studies are made on the Safety Risk Model (SRM) developed by the Rail Safety and Standards Board (RSSB). RSSB was established on 1 April 2003, as the implementation of one of the core recommendations from the second part of Lord Cullen's public inquiry into Ladbroke Grove train accident. The prime objective of RSSB is to lead and facilitate the railway industry's work to achieve continuous improvement in the safety performance on the Great Britain mainline railways. As a part of the efforts, RSSB has built the SRM to measure the risk and the underlying causes [Harrison 2004].

The RSSB-SRM is composed of a series of Fault-tree and Event tree models corresponding to the 125 hazardous events respectively. Currently the "best-estimated" values are set to the input parameters; consequently a point estimation of

the yearly equivalent expected fatalities can be obtained by running the model [Dennis 2006]. To support decision making, the assessment of the uncertainty in the output is needed. It therefore offers good case studies to test the procedure and the methods developed in the previous chapters. Our case studies are to be conducted on the hazardous events HET10 and HET12 as they are the two largest contributors to the safety risk in RSSB-SRM [Dennis 2006]. The acronym HET stands for Hazardous Event related to Train. More details on the two hazardous events are to be included later in the case studies.

6.2 Elicitation Process

In line with the processes summarized in [O'Hagan, Buck et al. 2006], we design the elicitation process for building up the uncertainty model of multiple input parameters. The major development of the process is on the development of the structures to model the dependency of high dimensional input, which is the most complex and difficult part of building uncertainty model of multiple input parameters [Kurowicka and Cooke 2006; O'Hagan, Buck et al. 2006].

As shown in Fig. 6.1, the whole elicitation process consists of 6 stages. At Stage 1, we develop models describing the uncertainty in the basic events of the fault trees and the uncertainty in the consequences. As developed in Chapter 3 and Chapter 4 respectively, the models require reasonable elicitation workload that makes the models practical for the stakeholders.

At Stage 2 we choose the experts for the elicitation. The experts are chosen from RSSB for three reasons. First, it is the RSSB that builds the safety risk model. The RSSB now is concerned with their uncertainty in the model. The experts are likely to have a positive attitude towards this exercise. Second, the experts know how the RSSB-SRM is built and how it is used in practice. As such, they are knowledgeable about the uncertainty sources. Third, the experts have at least basic understanding of statistics. Therefore they can understand elicitation questions.

At Stage 3, we train the experts with the procedures and models for modelling the uncertainties in the basic events and the consequences respectively. At the first step of the training, the experts read the documents on the models. After that, a meeting was organized on which the models are discussed and demonstration examples were made. After the training process, the experts understand the types of information to be elicited and how the elicited information is used to assess the uncertainty of the input parameters.

At Stage 4: we conduct the elicitation to build up the uncertainty models for the basic events and the consequences respectively. At the first step, we develop a software tool with MS Excel to build up the benchmark of the qualitative correlation assessment by the method described in Chapter 3. The details of the benchmark are included in Section 6.3.

For each case study, we elicit the experts' judgement to build the covariance matrix of the basic events of the fault tree. The inconsistent assessments regarding the correlations between the basic events and the uncertainty factors can be identified during the process by the method developed in Chapter 3. For each case study, we also build up the uncertainty model of the consequences of the event tree through elicitation. The details of the elicitation are included in the case studies.

At Stage 5: we build up the uncertainty models in the input parameters based on the elicited information. For each case study, we build up the covariance matrix of the basic events of the fault tree by the procedure developed in Chapter 3. We build up also the uncertainty model of the consequences by the method developed in Chapter 4.

At State 6 the whole elicitation work is reviewed. The experts check the elicited data and make sure that no important factors are missed. The experts also check the outcome correlations among the basic events and the plots of the samples of the rule sets. When the experts find any outcomes inconsistent with their judgement, the

associated assessments are then adjusted and the uncertainty models in the inputs are updated accordingly.

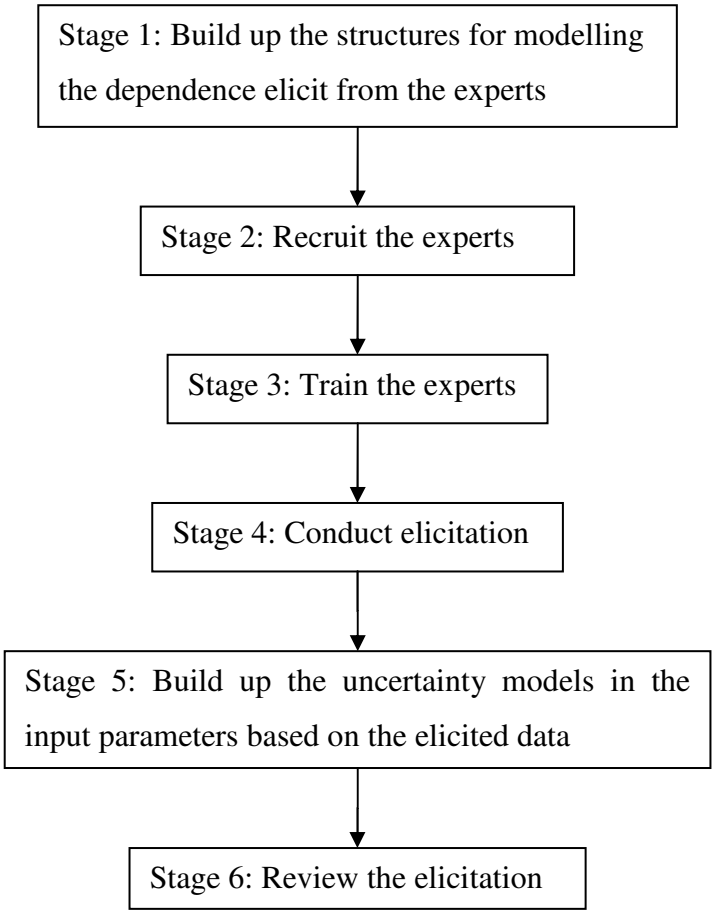


Fig. 6.1 The elicitation process for the case studies

6.3 Benchmark of the Qualitative Correlation Assessment

Following the methods developed in Chapter 4, qualitative assessment plus benchmark is used for the elicitation of the correlations. Based on the literature and the discussion with the experts, we use five correlation levels including: Very Weak (VW), Weak(W), Medium(M), Strong(S), and Very Strong(VS). The benchmark exercise is conducted with 200 samples of the preset correlation ρ_{XY}. These preset correlations are then pooled into five groups based on the expert assessment. The minimum, average and the maximum of the correlations associated with the five

levels are summarized in Table 6.1. The average values are to be used for the case studies.

Table 6.1 Summary of the statistics of the qualitative correlation levels

Correlation Levels	Minimum	Average	Maximum
VW	0.007322779	0.179473026	0.387266054
W	0.039863085	0.330255985	0.565863759
M	0.34533765	0.563091162	0.711791892
S	0.683938713	0.775600499	0.843917637
VS	0.848889018	0.939226752	0.999453956

6.4 Case Study with HET10

6.4.1 Elicitation process for building the covariance matrix of the precursors

The hazardous event HET10 is defined as *passenger train collision with road vehicle on level crossings* [Dennis 2005]. On the British mainline railways, there are eight types of level crossings (LC) called: ABCL, AHB, AOCL, MG/B, OC, UWC, UWC, MWL and UWC(T) [RSSB 2004]. The basic events of the fault tree are called precursor in the RSSB-SRM. There are 52 precursors related to HET10. Each precursor is associated with one type of level crossings. Therefore a group of precursors can be defined on the same event except being associated with the different types of level crossings. For instance, a group of precursors are defined as *road vehicle (RV) driver error causing RV struck by train on level crossing L,* where *L* denotes one of the eight types of LC. The experts believe that such a group of precursors form a family. The variant factors of the family are the uncertainty factors that are related to or affected by any differences among the eight types of level crossings. After the effect of the variant factors is accounted, the residual uncertainties of the family members form an invariant factor. Therefore the linear model of the family is defined as in Equation 3.80. The experts believe as well that there is no common factor between the families and, as a result, the precursors from different families are independent. Accordingly the elicitation process for building up

the covariance matrix of the precursors is made as shown in Fig. 6.2. The elicitation process starts with identifying the families by the definitions of the precursors. Secondly the differences among the different types of level crossings are elicited. Accordingly, the variant factors are identified for each precursor family.

For the elicited variant factor classes, the correlations between the precursors and the variant factors are elicited as shown in Block 4a. The explanation coefficient q_i as defined in Equation 3.71 is calculated for each member as shown in Block 6. The value of $q_i > 1$ indicates that the experts over estimate some variant factors and therefore need to adjust their judgement as shown in Block 7. When the coefficient q_i is very small for all the family members, the correlations within a factor class have insignificant impact on the family correlation matrix. Therefore they are assumed to be zero. Otherwise, the correlations within a factor class need to be elicited as shown in Block 5.

For some families related with HET10, no significant variant factors can be elicited. In this case, the proportion of the variance due to the family commonality, i.e. $1 - q_i$, is elicited as shown in Block 4b. The marginal variance is attributed to an unspecified exclusive factor for each member precursor as shown in Block 9.

With the elicitations regarding the variant factors, the family correlation matrix can be derived in Block 10. The variance of the family is elicited in Block 2 and the family covariance matrix can then be derived in Block 11 by the methods developed in Chapter 3.

We will continue to discuss elicitation in details. To make the elicitation easy to follow, the following subsections are headed according to the structure of Fig. 6.2. At the beginning of each subsection, we also locate the subsection with reference to the blocks in Fig. 6.2.

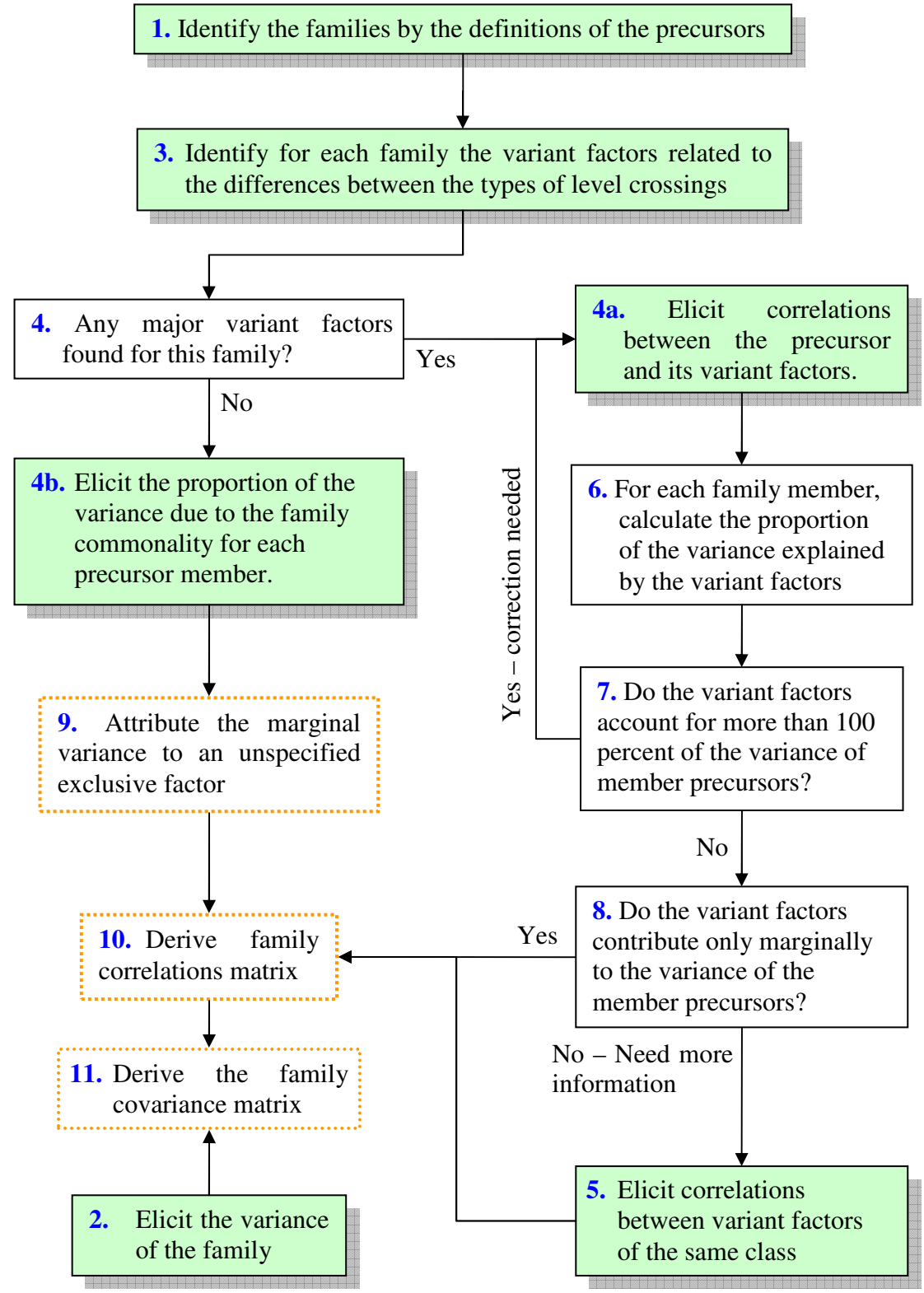


Fig. 6.2 Flow chart for building covariance matrix of the precursors

6.4.2 Precursor Families

This subsection is referred to Block 1 in Fig. 6.2. We will identify the precursor families.

The 52 precursors related to HET10 are put into nine groups by the definitions. The experts believe that the group of "*RV Driver error causing RV struck by train on various types of LCs"* and the group of "*RV driver deliberate action causing RV struck by train on various types of LCs"* are identical. The two groups are therefore combined. As a result, eight precursor families are identified as summarized in Table 6.2. Within each family, the precursors are defined on various types of level crossings. As summarized in Table 6.3, a tick indicates that the precursor family in the most left column has a precursor defined on the type of level crossing as titled on the top row.

Table 6.2 Precursor families identified for RSSB-SRM HET10

No.	Family Code	Description
1	RVSTRENV	RV incorrectly on various types of LCs and struck by train due to environmental factors
2	TOVRSPD	Train over-speeding causes RV struck by train on various types of LCs.
3	RVSTRAN	RV stranded and struck by train on the various types of LCs
4	SPADPROT	RV struck by train due to SPAD at signal protecting the various types of LCs
5	SIGERR	Signalman or Crossing keeper error causes RV struck by train on various types of LCs.
6	RVDRVERR	RV driver error causing RV struck by train on various types of LCs and; RV driver deliberate action causing RV struck by train on various types of LCs
7	RVDRVSUI	RV stuck by train due to RV driver suicide on various types of LCs.
8	LTBRFAI	RV stuck by train due to Lights/Barriers fail to operate on various types of LCs

Table 6.3 Precursor families and the associated types of level crossings

Family Code	ABCL	AHB	AOCL	MG/B	OC	UWC	UWC MWL	UWC (T)
1 RVSTRENV	√	√	√	√	√	√	√	√
2 TOVRSPD	√	√	√	√				

3 RVSTRAN	√	√	√	√		√	√	√
4 SPADPROT	√		√	√				
5 SIGERR		√		√				√
6 RVDRVERR	√	√	√	√	√	√	√	√
7 RVDRVSUI	√	√	√		√	√	√	√
8 LTBRFAI	√	√	√	√				

6.4.3 Variance of Families

This subsection is referred to Block 2 in Fig. 6.2. We will elicit the variance of each of the eight precursor families as identified in the subsection 6.4.2.

The occurrence rate of a family, denoted as λ, is the sum of the occurrence rates of the precursors in the family. It is assumed that λ follows normal distribution. The mean of λ is, denoted as μ, is set as the sum of the existing point values of the family members. Given λ, the waiting time between two successive occurrences, denoted as T, is assumed to be exponentially distributed. Subsequently, the predictive distribution of T can be developed by Bayes theorem. The waiting time T is an observable variable and the expert's ability in assessing the percentiles has been well proven [Pearson and Tukey 1965; Keefer and Bodily 1983; Cooke 1991; Garthwaite, Kadane et al. 2005; Kurowicka and Cooke 2006; O'Hagan, Buck et al. 2006]. As developed in Chapter 3, we will elicit the 95 percentile of the waiting time, denoted as $t_{0.95}$, and consequently derive σ_λ, i.e. the standard deviation of λ.

As summarized in Table 6.4, the value of $t_{0.95}$ is elicited within a reference interval for each family. The minimum $t_{0.95}$ is calculated on with the pure exponential distribution when λ is fixed to μ. The experts agree that σ_λ has a maximum boundary at $\mu/3.8$, based on with the possible maximum $t_{0.95}$ is calculated accordingly. The experts then are asked to give their belief of $t_{0.95}$ within the given interval. The elicited $t_{0.95}$ and the derived σ_λ for the eight families related to HET10 are presented in Table 6.4. These standard deviations are to be used in developing the family covariance matrix.

Table 6.4 Elicited 95 percentile of the waiting time, i.e. $t_{0.95}$, and the outcome standard deviation of the occurrence rate of the families related with HET10

Family Code	Minimum $t_{0.95}$	Maximum $t_{0.95}$	Assessment of $t_{0.95}$	Standard Deviation σ_λ
RVSTRENV	111 months	126 months	118 months	2.36E-10
TOVRSPD	93 years	106 years	100 years	2.33E-11
RVSTRAN	42 months	47 months	44 months	6.28E-10
SPADPROT	197 years	224 years	210 years	1.10E-11
SIGERR	110 months	125 months	118 months	2.36E-10
RVDRVERR	78 days	89 days	83 days	1.02E-08
RVDRVSUI	70 months	80 months	75 months	3.71E-10
LTBRFAI	93 years	106 years	100 years	2.33E-11

Based on the variance elicited above, we are going to derive the covariance matrix of each precursor family. For this purpose, we need to elicit for each family the variant factors and the information on the variant factors. The elicitation starts from Block 3 in Fig. 6.2. The choice of the following route depends on the answer to Block 4 in Fig. 6.2. When there is a variant factor, we follow the route through Block 4a, Block 5, Block 10 to Block 11. When there is no variant factor, we follow the route through Block 4b, Block 10 to Block 11. Each of the following eight subsections is dedicated to one of the eight precursor families summarized as Table 6.2.

6.4.4 Covariance Matrix of the Precursor Family RVSTRENV

6.4.4.1 Variant Factors

This subsection is related to Block 3 in Fig. 6.2. We will elicit the variant factors for the precursor family RVSTRENV.

The precursor family RVSTRENV is defined for *"RV incorrectly on various types of LCs and struck by train due to environmental factors"*. As summarized in Table 6.5, one variant factor class is elicited as *"RV drivers' ability to respond to the prevailing*

weather conditions". The level crossings are put into two categories: user worked crossings (UWC) and non-UWC [RSSB 2004]. The two categories of LCs have their own users. Accordingly the two groups of users are defined as UWC users and non-UWC users.

Table 6.5 Elicited factor class for the family RVSTRENV

Family	RVSTRENV – RV incorrectly on various types of LCs and struck by train due to environmental factors
No.	**Variant Class**
1.	RV drivers' ability to respond to the prevailing weather conditions.

6.4.4.2 Correlations between the Precursors and the Associated Variant Factors

This subsection is related to Block 4a in Fig. 6.2 because one variant factor class has been elicited for the family RVSTRENV. We will elicit the correlations between the precursors and the associated variant factors.

A medium correlation is elicited between the precursors on UWC LCs and the UWC users; a medium correlation is also elicited between the precursors on non-UWC LCs and the non-UWC users. The elicited results are summarized in Table 6.6. Because the elicited correlations are positive in this case study we do not mark the correlation sign explicitly.

Table 6.6 Elicited correlations between the variant factors and the associated precursors of the family RVSTRENV

Precursor Family		RVSTRENV – RV incorrectly on LC and struck by train due to environmental factors	
Variant Class		RV drivers' ability to respond to the prevailing weather conditions.	
NO.	**Variant Factors**	**Associated precursor**	**Correlation**
1.	Concerning UWC LCs	RVSTRENV on UWC LCs	M
2.	Concerning Non-UWC LCs	RVSTRENV on Non-UWC LCs	M

6.4.4.3 Correlations within the Variant Factor Class

This subsection is related to Block 5 in Fig. 6.2. We will elicit the correlations within the factor class of *"RV drivers' ability to respond to the prevailing weather conditions"*.

A medium correlation between UWC users and non-UWC users is elicited as summarized in Table 6.7.

Table 6.7 Correlations within the factor class for the family RVSTRENV

Family name	RVSTRENV – RV incorrectly on LCs and struck by train due to environmental factors
Variant class	RV drivers' ability to respond to the prevailing weather conditions
Variant factor 1	On UWC LCs
Variant factor 2	On Non-UWC LCs
The correlation between the above two variant factors	M

6.4.4.4 Family Correlation Matrix

This subsection is related to Block 10 in Fig. 6.2. We will derive the family correlation matrix based on the elicitations on the variant factors.

The above qualitative correlations regarding the variant factors are then mapped to the benchmark averages as summarized in Table 6.1. The correlation matrix of the family RVSTRENV can then be derived. As summarized in Table 6.8, the family correlation is divided into two blocks corresponding to the UWC users group and the non-UWC users group.

Table 6.8 Correlation matrix of the family RVSTRENV

	ABCL	AHB	AOCL	MG/B	OC	UWC	UWC-MWL	UWC+T
ABCL	1	1	1	1	1	0.86	0.86	0.86
AHB	1	1	1	1	1	0.86	0.86	0.86
AOCL	1	1	1	1	1	0.86	0.86	0.86
MG/B	1	1	1	1	1	0.86	0.86	0.86

OC	1	1	1	1	1	0.86	0.86	0.86
UWC	0.86	0.86	0.86	0.86	0.86	1	1	1
UWC-MWL	0.86	0.86	0.86	0.86	0.86	1	1	1
UWC+T	0.86	0.86	0.86	0.86	0.86	1	1	1

6.4.4.5 Family Covariance Matrix

This subsection is related to Block 11 in Fig. 6.2. We will derive the family covariance matrix.

Based one the family variance as summarized in Table 6.4 and the above elicitations on the variant factors, the family covariance matrix is derived by the method developed in Chapter 3. The outcome family covariance matrix is summarized in Table 6.9.

Table 6.9 Covariance matrix of the precursor family RVSTRENV

WABCL ENVTE	WAHB-ENVTE	WAOCL ENVTE	WMG/B ENVTR	WOC--ENVTE	WUWC-ENVTE	WUWCM ENVTE	WUWCT ENVTE
5.48E-24	4.73E-23	1.41E-23	8.69E-23	6.04E-24	1.99E-22	1.34E-23	1.4E-22
4.73E-23	4.08E-22	1.22E-22	7.5E-22	5.21E-23	1.71E-21	1.16E-22	1.21E-21
1.41E-23	1.22E-22	3.62E-23	2.24E-22	1.55E-23	5.1E-22	3.45E-23	3.6E-22
8.69E-23	7.5E-22	2.24E-22	1.38E-21	9.58E-23	3.15E-21	2.13E-22	2.22E-21
6.04E-24	5.21E-23	1.55E-23	9.58E-23	6.65E-24	2.19E-22	1.48E-23	1.54E-22
1.99E-22	1.71E-21	5.1E-22	3.15E-21	2.19E-22	9.69E-21	6.56E-22	6.84E-21
1.34E-23	1.16E-22	3.45E-23	2.13E-22	1.48E-23	6.56E-22	4.44E-23	4.63E-22
1.4E-22	1.21E-21	3.6E-22	2.22E-21	1.54E-22	6.84E-21	4.63E-22	4.83E-21

6.4.5 Precursor Family TOVRSPD

The precursor family TOVRSPD is defined as *"Train over-speeding causes RV struck by train on various types of LCs"*. For this family, no significant variant factor is identified by the experts. As proposed previously, the proportion of the variance due to the family commonality is elicited for each family member as presented in

Appendix B. The family correlation matrix and the family covariance matrix are derived as presented in Table 6.10 and Table 6.11 respectively.

Table 6.10 Correlation matrix of the precursor family TOVRSPD

	ABCL	AHB	AOCL	MG/B
ABCL	1	0.9	0.9	0.9
AHB	0.9	1	0.9	0.9
AOCL	0.9	0.9	1	0.9
MG/B	0.9	0.9	0.9	1

Table 6.11 Covariance matrix of the precursor family TOVRSPD

PABCLOVSTH	PAHB-OVSTH	PAOCLOVSTH	PMG/BOVSTH
5.61E-25	5E-24	1.49E-24	9.2E-24
5E-24	5.51E-23	1.48E-23	9.11E-23
1.49E-24	1.48E-23	4.89E-24	2.71E-23
9.2E-24	9.11E-23	2.71E-23	1.86E-22

6.4.6 Precursor Family RVSTRAN

The precursor family RVSTRAN is defined as *"RV stranded on LC causes RV struck by train on various types of LCs"*. For this family, two variant factor classes are identified as summarized as:

- FC1: Propensity for there to be blocking back on a crossing, i.e. where you get traffic jams extending back over level crossings
- FC2: Profile of the RV drivers using the crossing, i.e. the propensity of certain drivers to violate rules

The elicited data regarding the two variant factors are presented in Appendix B. The family correlation matrix and the family covariance matrix are derived as presented in Table 6.12 and Table 6.13.

Table 6.12 Correlation matrix of the precursor family RVSTRAN

	ABCL	AHB	AOCL	MG/B	UWC	UWC-MWL	UWC+T
ABCL	1.00	0.62	0.86	0.85	0.72	0.72	0.72

AHB	0.62	1.00	0.41	0.40	0.46	0.46	0.46
AOCL	0.86	0.41	1.00	0.94	0.65	0.65	0.65
MG/B	0.85	0.40	0.94	1.00	0.67	0.67	0.67
UWC	0.72	0.46	0.65	0.67	1.00	1.00	1.00
UWC-MWL	0.72	0.46	0.65	0.67	1.00	1.00	1.00
UWC+T	0.72	0.46	0.65	0.67	1.00	1.00	1.00

Table 6.13 Covariance matrix of the precursor family RVSTRAN

VABCLSTRTE	VAHB-STRTE	VAOCLSTRTE	VMG/BSTRTE	VUWC-STRTE	VUWCMSTRTE	VUWCTSTRTE
5.39E-23	2.64E-22	1.1E-22	6.69E-22	1.5E-21	1.02E-22	1.06E-21
2.64E-22	3.38E-21	4.14E-22	2.49E-21	7.64E-21	5.17E-22	5.39E-21
1.1E-22	4.14E-22	2.99E-22	1.74E-21	3.18E-21	2.15E-22	2.25E-21
6.69E-22	2.49E-21	1.74E-21	1.14E-20	2.02E-20	1.37E-21	1.43E-20
1.5E-21	7.64E-21	3.18E-21	2.02E-20	8.01E-20	5.41E-21	5.65E-20
1.02E-22	5.17E-22	2.15E-22	1.37E-21	5.41E-21	3.67E-22	3.82E-21
1.06E-21	5.39E-21	2.25E-21	1.43E-20	5.65E-20	3.82E-21	4E-20

6.4.7 Precursor Family SPADPROT

The precursor family SPADPROT is defined as *"SPAD at signal protecting the LC causes RV struck by train on various types of LCs"*. One factor class is identified by the experts as "the effectiveness of the signal protecting the LCs". The elicited data regarding the variant factors are presented in Appendix B. The family correlation matrix and the family covariance matrix are derived as presented in Table 6.14 and Table 6.15.

Table 6.14 Correlation matrix of the precursor family SPADPROT

	ABCL	AOCL	MG/B
ABCL	1.00	0.97	0.97
AOCL	0.97	1.00	0.97
MG/B	0.97	0.97	1.00

Table 6.15 Covariance matrix of the precursor family SPADPROT

PABCLSPDTH	PAOCLSPDTH	PMG/BSPDTH

2.34E-25	6.96E-25	4.3E-24
6.96E-25	2.19E-24	1.31E-23
4.3E-24	1.31E-23	8.34E-23

6.4.8 Precursor Family SIGERR

The precursor family SIGERR is defined as *"Signalman or Crossing keeper error causes RV struck by train on various types of LCs"*. One factor class is identified as *"Complexity of operating the crossing"*. The elicited data regarding the variant factors are presented in Appendix B. The family correlation matrix and the family covariance matrix are derived as presented in Table 6.16 and Table 6.17.

Table 6.16 Correlation matrix of the precursor family SIGERR

	AHB	MG/B	UWC+T
AHB	1.00	0.74	0.93
MG/B	0.74	1.00	0.74
UWC+T	0.93	0.74	1.00

Table 6.17 Covariance matrix of the precursor family SIGERR

LAHB-SKPTH	LMG/BSKPTH	VUWCTSKPTH
2.2E-21	2.47E-21	5.8E-21
2.47E-21	5.07E-21	7.02E-21
5.8E-21	7.02E-21	1.78E-20

6.4.9 Precursor Family RVDRVERR and RVDRVDEL

The experts believe that the precursors *"RV Driver error causing RV struck by train on LCs"* and the precursors *"RV driver deliberate action on various types of crossing"* are identical on the same level crossing in terms of the uncertainty modelling. Correspondingly these two groups of precursors are defined within in one family RVDRVERR and RVDRVDEL. One factor class is identified by the experts as *"the quality of the users"*. The elicited data regarding the variant factors are presented in Appendix B.

The correlation matrix for the family RVDRVERR and RVDRVDEL is then derived as presented in Table 6.18. The correlation matrix composes eight rows and eight columns corresponding to the eight LCs because the RV Driver error and the RV driver deliberate action on the same LC are combined together. Denote this 8×8 correlation matrix as r_c , where the subscript c stands for combination. The correlation matrix for the 16 precursors related to RVDRVERR and RVDRVDEL respectively can then be defined as

$$r(\mathrm{ERR}, \mathrm{EDL}) = \begin{bmatrix} r_c & r_c \\ r_c & r_c \end{bmatrix}$$

Table 6.18 Correlation matrix of the precursor family RVDRVERR and RVDRVDEL

	ABCL	AHB	AOCL	MG/B	OC	UWC	UWC-MWL	UWC+T
ABCL	1.00	1.00	1.00	1.00	1.00	0.84	0.84	0.84
AHB	1.00	1.00	1.00	1.00	1.00	0.84	0.84	0.84
AOCL	1.00	1.00	1.00	1.00	1.00	0.84	0.84	0.84
MG/B	1.00	1.00	1.00	1.00	1.00	0.84	0.84	0.84
OC	1.00	1.00	1.00	1.00	1.00	0.84	0.84	0.84
UWC	0.84	0.84	0.84	0.84	0.84	1.00	1.00	1.00
UWC-MWL	0.84	0.84	0.84	0.84	0.84	1.00	1.00	1.00
UWC+T	0.84	0.84	0.84	0.84	0.84	1.00	1.00	1.00

Denote the covariance matrix for the combined family as:

$$\boldsymbol{\Sigma}_c = \left[\sigma_{i,j}^2\right],\ i, j = 1, \cdots, 8 \tag{6.1}$$

where c stands for the combination of RVDRVERR and RVDRVDEL.

The covariance matrix $\boldsymbol{\Sigma}_c$ can be derived as presented in Table 6.19. From $\boldsymbol{\Sigma}_c$ the covariance matrix for the 16 precursors related to RVDRVERR and RVDRVDEL can be defined as

$$\operatorname{cov}(\mathrm{ERR},\mathrm{DEL})=\left[\begin{array}{c|c}\operatorname{cov}(\mathrm{ERR}) & \operatorname{cov}(\mathrm{ERR},\mathrm{DEL})\\ \hline \operatorname{cov}(\mathrm{DEL},\mathrm{ERR}) & \operatorname{cov}(\mathrm{DEL})\end{array}\right] \tag{6.2}$$

where $\operatorname{cov}(\mathrm{ERR})=\left[E_{i,j}^2\right]$ and $\operatorname{cov}(\mathrm{DEL})=\left[D_{i,j}^2\right]$ are the covariance matrix related to RVDRVERR and RVDRVDEL respectively; $\operatorname{cov}(\mathrm{ERR},\mathrm{DEL})=\left[ED_{i,j}^2\right]$ is the covariance matrix between RVDRVERR and RVDRVDEL; $\operatorname{cov}(\mathrm{DEL},\mathrm{ERR})$ is the transposition of $\operatorname{cov}(\mathrm{ERR},\mathrm{DEL})$.

Based on Equation 6.2, we are going to derive $E_{i,j}^2$, $D_{i,j}^2$ and $ED_{i,j}^2$. Because the RV Driver error and the RV driver deliberate action on the same LC are believed identical, they can be defined by the same random variable with different factors. The factors are defined by the associated means as summarized in Table 6.20. We define

$$c_{E,i}=\frac{\mu_{E,i}}{\mu_{E,i}+\mu_{D,i}} \tag{6.3}$$

$$c_{D,i}=\frac{\mu_{D,i}}{\mu_{E,i}+\mu_{D,i}} \tag{6.4}$$

where $\mu_{E,i}$ and $\mu_{D,i}$ stand for are the mean of the i^{th} precursor corresponding to RVDRVERR and RVDRVDEL respectively.

Based on Equations 6.3 and 6.4, we have

$$E_{i,j}^2=c_{E,i}\cdot c_{E,j}\cdot\sigma_{i,j}^2 \tag{6.5}$$

$$D_{i,j}^2=c_{D,i}\cdot c_{D,j}\cdot\sigma_{i,j}^2 \tag{6.6}$$

$$ED_{i,j}^2=c_{E,i}\cdot c_{D,j}\cdot\sigma_{i,j}^2 \tag{6.7}$$

By Equations 6.5-6.7, the matrix $\mathrm{cov}(\mathrm{ERR})$, $\mathrm{cov}(\mathrm{DEL})$ and $\mathrm{cov}(\mathrm{ERR},\mathrm{DEL})$ can be calculated from Σ_c and are presented in Tables 6.21-6.23 respectively.

Table 6.19 Covariance matrix of the precursor family RVDRVERR and RVDRVDEL

ABCL	AHB	AOCL	MG/B	OC	UWC	UWC-MWL	UWC+T
9.20E-21	8.38E-20	2.5E-20	1.54E-19	1.07E-20	3.44E-19	2.33E-20	2.43E-19
8.38E-20	7.64E-19	2.28E-19	1.4E-18	9.76E-20	3.14E-18	2.12E-19	2.22E-18
2.50E-20	2.28E-19	6.78E-20	4.19E-19	2.91E-20	9.35E-19	6.33E-20	6.6E-19
1.54E-19	1.4E-18	4.19E-19	2.58E-18	1.79E-19	5.77E-18	3.91E-19	4.07E-18
1.07E-20	9.76E-20	2.91E-20	1.79E-19	1.25E-20	4.01E-19	2.71E-20	2.83E-19
3.44E-19	3.14E-18	9.35E-19	5.77E-18	4.01E-19	1.81E-17	1.23E-18	1.28E-17
2.33E-20	2.12E-19	6.33E-20	3.91E-19	2.71E-20	1.23E-18	8.31E-20	8.67E-19
2.43E-19	2.22E-18	6.6E-19	4.07E-18	2.83E-19	1.28E-17	8.67E-19	9.05E-18

Table 6.20 Means of the precursors related to RVDRVERR and RVDRVDEL

	ABCL	AHB	AOCL	MG/B	OC	UWC	UWC-MWL	UWC+T
ERR	2.42E-10	7.45E-09	1.33E-08	3.92E-10	1.96E-09	5.10E-09	2.75E-09	1.02E-08
DEL	8.97E-11	2.66E-09	3.53E-09	3.92E-10	2.28E-14	1.17E-09	2.38E-10	1.34E-09

Table 6.21 Covariance matrix related to RVDRVERR

VABCLDRRTH	VAHB-DRRTH	VAOCLDRRTH	VMG/BDRRTH	VOC--DRRTH	VUWC-DRRTH	VUWCMDRRTH	VUWCTDRRTH
4.89E-21	4.51E-20	1.44E-20	5.62E-20	7.81E-21	2.04E-19	1.56E-20	1.57E-19
4.51E-20	4.15E-19	1.33E-19	5.18E-19	7.19E-20	1.88E-18	1.44E-19	1.44E-18
1.44E-20	1.33E-19	4.24E-20	1.65E-19	2.3E-20	6.01E-19	4.6E-20	4.61E-19
5.62E-20	5.18E-19	1.65E-19	6.46E-19	8.97E-20	2.35E-18	1.8E-19	1.8E-18
7.81E-21	7.19E-20	2.3E-20	8.97E-20	1.25E-20	3.26E-19	2.5E-20	2.5E-19
2.04E-19	1.88E-18	6.01E-19	2.35E-18	3.26E-19	1.2E-17	9.19E-19	9.21E-18
1.56E-20	1.44E-19	4.6E-20	1.8E-19	2.5E-20	9.19E-19	7.04E-20	7.06E-19
1.57E-19	1.44E-18	4.61E-19	1.8E-18	2.5E-19	9.21E-18	7.06E-19	7.07E-18

Table 6.22 Covariance matrix related to RVDRVDEL

VABCLDELTH	VAHB-DELTH	VAOCLDELTH	VMG/BDELTH	VOC--DELTH	VUWC-DELTH	VUWCMDELTH	VUWCTDELTH
6.73E-22	5.96E-21	1.41E-21	2.09E-20	3.37E-26	1.74E-20	5.02E-22	7.62E-21

5.96E-21	5.28E-20	1.25E-20	1.85E-19	2.98E-25	1.54E-19	4.45E-21	6.75E-20
1.41E-21	1.25E-20	2.97E-21	4.38E-20	7.08E-26	3.66E-20	1.05E-21	1.6E-20
2.09E-20	1.85E-19	4.38E-20	6.46E-19	1.04E-24	5.39E-19	1.55E-20	2.36E-19
3.37E-26	2.98E-25	7.08E-26	1.04E-24	1.69E-30	8.71E-25	2.51E-26	3.81E-25
1.74E-20	1.54E-19	3.66E-20	5.39E-19	8.71E-25	6.33E-19	1.83E-20	2.77E-19
5.02E-22	4.45E-21	1.05E-21	1.55E-20	2.51E-26	1.83E-20	5.27E-22	8E-21
7.62E-21	6.75E-20	1.6E-20	2.36E-19	3.81E-25	2.77E-19	8E-21	1.21E-19

Table 6.23 Covariance between RVDRVERR and RVDRVDEL

	VABCLDELTH	VAHB-DELTH	VAOCLDELTH	VMG/BDELTH	VOC--DELTH	VUWC-DELTH	VUWCMDELTH	VUWCTDELTH
VABCLDRRTH	1.82E-21	1.61E-20	3.81E-21	5.62E-20	9.08E-26	4.69E-20	1.35E-21	2.05E-20
VAHB-DRRTH	1.67E-20	1.48E-19	3.51E-20	5.18E-19	8.37E-25	4.32E-19	1.25E-20	1.89E-19
VAOCLDRRTH	5.34E-21	4.73E-20	1.12E-20	1.65E-19	2.67E-25	1.38E-19	3.98E-21	6.05E-20
VMG/BDRRTH	2.09E-20	1.85E-19	4.38E-20	6.46E-19	1.04E-24	5.39E-19	1.55E-20	2.36E-19
VOC—DRRTH	2.9E-21	2.56E-20	6.08E-21	8.97E-20	1.45E-25	7.48E-20	2.16E-21	3.28E-20
VUWC-DRRTH	7.58E-20	6.71E-19	1.59E-19	2.35E-18	3.79E-24	2.76E-18	7.95E-20	1.21E-18
VUWCMDRRTH	5.8E-21	5.14E-20	1.22E-20	1.8E-19	2.9E-25	2.11E-19	6.09E-21	9.25E-20
VUWCTDRRTH	5.82E-20	5.15E-19	1.22E-19	1.8E-18	2.91E-24	2.12E-18	6.11E-20	9.27E-19

6.4.10 Precursor Family RVDRVSUI

The precursor family RVDRVSUI is defined as *"RV driver suicide causes RV struck by train on various types of LCs"*. For this family, no significant variant class is identified by the experts. As proposed previously, the proportion of the variance due to the family commonality is elicited for each precursor as presented in Appendix B. The family correlation matrix and the family covariance matrix are derived as presented in Table 6.24 and Table 6.25.

Table 6.24 Correlation of the precursor family RVDRVSUI

	ABCL	AHB	AOCL	OC	UWC	UWC-MWL	UWC+T
ABCL	1	0.95	0.95	0.95	0.95	0.95	0.95

AHB	0.95	1	0.95	0.95	0.95	0.95	0.95
AOCL	0.95	0.95	1	0.95	0.95	0.95	0.95
OC	0.95	0.95	0.95	1	0.95	0.95	0.95
UWC	0.95	0.95	0.95	0.95	1	0.95	0.95
UWC-MWL	0.95	0.95	0.95	0.95	0.95	1	0.95
UWC+T	0.95	0.95	0.95	0.95	0.95	0.95	1

Table 6.25 Covariance matrix of the precursor family RVDRVSUI

VABCLSUIXPE	VAHBSUIXPE	VAOCLSUIXPE	VOC--SUIXPE	VUWCSUIXPE	VUWCLSUIXPE	VUWCTSUIXPE
1.32E-23	1.28E-22	3.82E-23	1.64E-23	6.24E-22	4.23E-23	4.41E-22
1.28E-22	1.37E-21	3.89E-22	1.67E-22	6.36E-21	4.3E-22	4.49E-21
3.82E-23	3.89E-22	1.22E-22	4.96E-23	1.89E-21	1.28E-22	1.34E-21
1.64E-23	1.67E-22	4.96E-23	2.24E-23	8.11E-22	5.49E-23	5.73E-22
6.24E-22	6.36E-21	1.89E-21	8.11E-22	3.26E-20	2.1E-21	2.19E-20
4.23E-23	4.3E-22	1.28E-22	5.49E-23	2.1E-21	1.49E-22	1.48E-21
4.41E-22	4.49E-21	1.34E-21	5.73E-22	2.19E-20	1.48E-21	1.63E-20

6.4.11 Precursor Family LTBRFAI

The precursor family LTBRFAI is defined as *"That Light/barriers fail to operate causes RV struck by train on various types of LCs"*. One factor class is identified by the experts as *"the technical workings of the lights and barriers and their failure rates"*. The elicited data regarding the variant factors are presented in Appendix B. The family correlation matrix and the family covariance matrix are derived as presented in Table 6.26 and Table 6.27.

Table 6.26 Correlation matrix of the precursor family LTBRFAI

	ABCL	AHB	AOCL	MG/B
ABCL	1.00	0.98	0.79	0.88
AHB	0.98	1.00	0.79	0.88
AOCL	0.79	0.79	1.00	0.88
MG/B	0.88	0.88	0.88	1.00

Table 6.27 Covariance matrix of the precursor family LTBRFAI

LABCLBLETF	LAHB-BLETF	LAOCLLSETF	LMG/BLSETF
7.43E-25	6.29E-24	1.5E-24	1.04E-23
6.29E-24	5.54E-23	1.3E-23	9E-23
1.5E-24	1.3E-23	4.91E-24	2.68E-23
1.04E-23	9E-23	2.68E-23	1.87E-22

6.4.12 Uncertainty Models in the Rule Sets

6.4.12.1 Identification of hazards

As summarized in Appendix B, there are 12 rule sets defined for the consequences related to HET10 [Dennis 2005]. Three hazard sources are identified as mechanical impact, fire, and toxic goods. For the mechanical impact, the train approaching speed is the major factor. The trains' approaching speeds are put into two levels associated with the automotive LCs (ABCL and AOCL) and the other level crossings respectively [RSSB 2004]. Accordingly two hazards are defined related to the mechanical impact as summarized in Table 6.28.

The fire hazard is affected by whether there are extra flammable goods involved in the accidence. Toxic goods are believed flammable and therefore cause extra casualties due to fire. As a result, three hazards related to fire are defined as fire without extra flammable goods, fire with flammable goods carried in the involved road vehicle, and fire with toxic goods as a flammable goods carried in the involved road vehicle as summarized in Table 6.28.

The toxic hazard source is specified for the toxic goods carried in the involved road vehicle only. Consequently one hazard related to toxic is defined as summarized in Table 6.28.

The five hazards identified above are to be defined on the injury atoms next.

Table 6.28 Hazards identified for HET10

Hazard Code	Hazard Source	Description
MIPAU	Mechanical impact	Mechanical impact with train approaching speed on automotive level crossings
MIPRO	Mechanical impact	Mechanical impact with train approaching speed on the other non-automotive level crossings
Fire	Fire	Fire without extra flammable goods or toxic goods carried in the involved road vehicle
FireFGs	Fire	Fire with extra flammable goods carried in the involved road vehicle
FireTGs	Fire	Fire with toxic goods as a flammable goods carried in the involved road vehicle
ToxicGs	Toxic goods	Toxic goods carried in the involved road vehicle

6.4.12.2 Definitions of Injury Atoms

The injury atoms are defined for each hazard source at the worst level. The injury atom associated with the mechanical impact is defined on the hazard associated with non-automotive level crossings. The injury atom associated with fire is defined on the hazard associated with extra flammable goods. The injury atom associated with toxic is defined on the only hazard with toxic goods carried in the road vehicle.

The injury atoms are assumed to follow Beta distributions that are derived by the methods developed in Chapter 4. As summarized in Table 6.29, the means and the 99 percentiles of the three reference injury atoms can be derived from the description of RSSB-SRM [Dennis 2005]. Accordingly the definition parameters of the Beta distribution, i.e. α and β, are derived for the injury atoms as presented in Table 6.29. The standard deviation is also included in Table 6.29. The hazards are then defined relative to the associated injury atoms as summarized in Table 6.30.

Table 6.29 Definitions of the injury atoms related to HET10

IA Code	The Mean	99 percentile	α	β	Std
MIPRO	7.543%	15.086%	7.09	86.93	0.027
FireFGs	14.2857%	28.571%	6.27	37.62	0.052

ToxicGs	21.428%	42.856%	5.38	19.74	0.0803

Table 6.30 Definitions of the hazards on the associated injury atoms related to HET10

	MIPRO (7.543%)	FireFGs (14.2857%)
MIPAU (1.103%)	0.1462	n/a
Fire (2.857%)	n/a	0.2
FireTGs (7.14285%)	n/a	0.5

6.4.12.3 Definitions of Rule Sets

The 12 rule sets related to HET10 are defined by the sequences of hazards as summarized in Table 6.31. A number 1 indicates that the rule set on the left column includes the hazard titled on the top row. Based on Table 6.30, the rule sets are then defined on the three injury atoms as summarized in Table 6.32. The 12 rule sets can be simulated by sampling the three injury atoms. Demonstrated in Fig. 6.3 are the smoothed density probability curves of the 6 rule sets related to the automotive level crossings. By Table 6.32, the other 6 rule sets have the similar patterns of the injury atoms and should have the similar shapes as demonstrated in Fig. 6.3 correspondingly.

Table 6.31 Definition of the rule sets related to HET10 on the hazards

Rule Sets	MIPAU	MIPRO	Fire	FireFGs	FireTGs	ToxicGs
T10-LCPAU-2	0	0	1	0	0	0
T10-LCPAU-3	1	0	0	0	0	0
T10-LCPAU-4	1	0	1	0	0	0
T10-LCPAU-5	0	0	0	0	0	1
T10-LCPAU-6	0	0	0	0	1	1
T10-LCPAU-7	0	0	0	1	0	0
T10-LCPRO-2	0	0	1	0	0	0
T10-LCPRO-3	0	1	0	0	0	0
T10-LCPRO-4	0	1	1	0	0	0
T10-LCPRO-5	0	0	0	0	0	1
T10-LCPRO-6	0	0	0	0	1	1
T10-LCPRO-7	0	0	0	1	0	0

Table 6.32 Definitions of the rule sets related to HET10 on the injury atoms

Rule Sets	MIPRO	FireFGs	ToxicGs	The mean
T10-LCPAU-2	0	0.2	0	0.02857
T10-LCPAU-3	0.1462	0	0	0.01103
T10-LCPAU-4	0.1462	0.2	0	0.03929
T10-LCPAU-5	0	0	1	0.21429
T10-LCPAU-6	0	0.5	1	0.27041
T10-LCPAU-7	0	1	0	0.14286
T10-LCPRO-2	0	0.2	0	0.02857
T10-LCPRO-3	1	0	0	0.07543
T10-LCPRO-4	1	0.2	0	0.10185
T10-LCPRO-5	0	0	1	0.21429
T10-LCPRO-6	0	0.5	1	0.27041
T10-LCPRO-7	0	1	0	0.14286

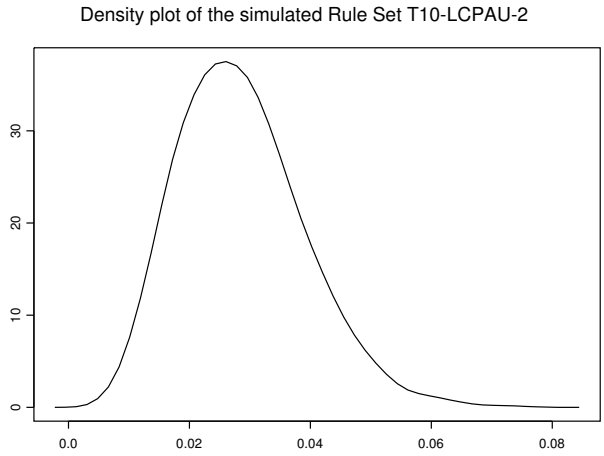


Fig. 6.3.a Density probability curve of the rule set T10-LCPAU-2

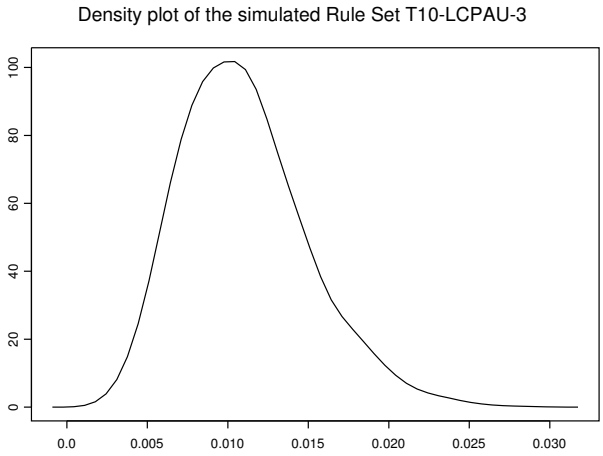


Fig. 6.3.b Density probability curve of the rule set T10-LCPAU-3

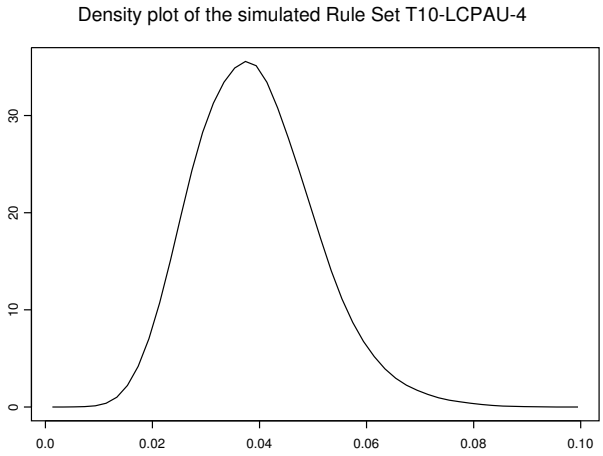


Fig. 6.3.c Density probability curve of the rule set T10-LCPAU-4

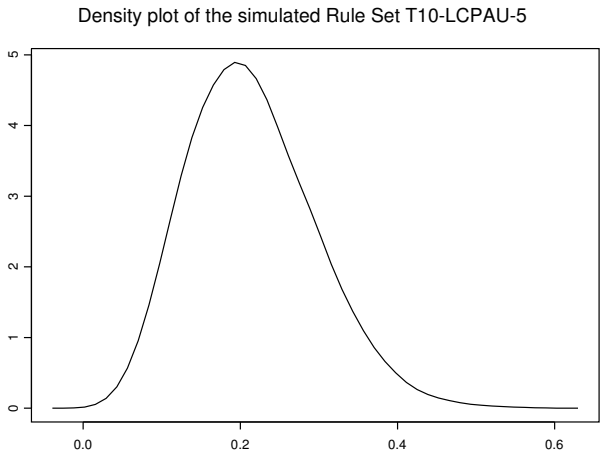


Fig. 6.3.d Density probability curve of the rule set T10-LCPAU-5

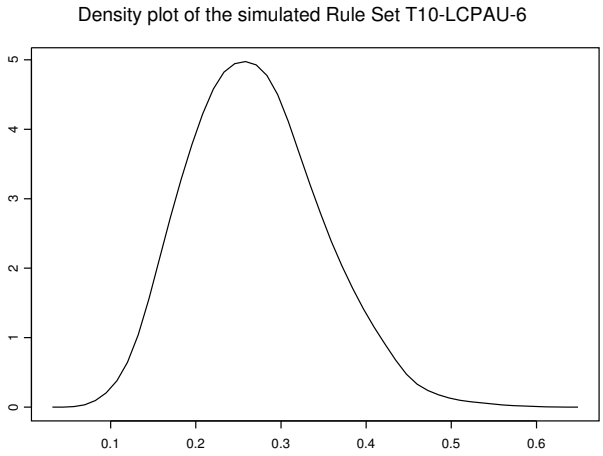


Fig. 6.3.e Density probability curve of the rule set T10-LCPAU-6

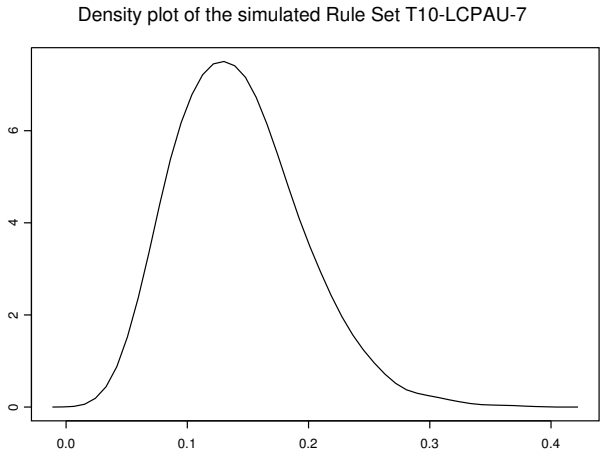


Fig. 6.3.f Density probability curve of the rule set T10-LCPAU-7

The scatter plots are then drawn to demonstrate the dependence between the rule sets. Shown Fig. 6.4 is the scatter plot between the rule sets T10-LCPPRO-4 and T10-LCPPRO-7. By Table 6.31, T10-LCPPRO-4 is defined by the sequence of the hazards MIPRO and Fire; T10-LCPPRO-7 is defined by the hazard FireFGs. The two rule sets are connected by the fire-related hazards Fire and ToxicFGs. The hazard MIPRO is however more than twice stronger than the hazard Fire in T10-LCPPRO-4. Consequently a moderate dependence is shown for the two rule sets.

Shown in Fig. 6.5 is the scatter plot between T10-LCPPRO-4 and T10-LCPPRO-3. By Table 6.31, the two rule sets share the hazard MIPRO. T10-LCPPRO-4 has the

extra hazard Fire. By Table 6.29 and Table 6.30, the hazard MIPRO is more than twice stronger than the hazard Fire. As a result, a much stronger dependence is shown for the two rule sets than that for T10-LCPPRO-4 and T10-LCPPRO-7 as shown in Fig. 6.4. Due to the extra hazard Fire, T10-LCPPRO-4 always has a higher individual casualty probability than T10-LCPPRO-3. This order relationship is demonstrated in Fig. 6.5 where all the samples fall under the 45-degree line.

Shown in Fig. 6.6 is the scatter plot between T10-LCPPRO-5 and T10-LCPAU-6. Based on Table 6.31, the two rule sets share the toxic-related hazards ToxicGs. The rule set T10-LCPPRO-6 has the extra hazard FireTGs. By Table 6.29 and Table 6.30, the hazard ToxicGs is three times stronger than the injury atom FireTGs. As a result, the stronger dependence is shown for the two rule sets than that for T10-LCPPRO-4 and T10-LCPPRO-3 as shown in Fig. 6.5. Due to the extra hazard FireTGs, the rule set T10-LCPPRO-6 has always larger individual casualty probability than the rule set T10-LCPPRO-5 does. This order relationship is demonstrated in Fig. 6.6 where all the samples fall above the 45-degree line.

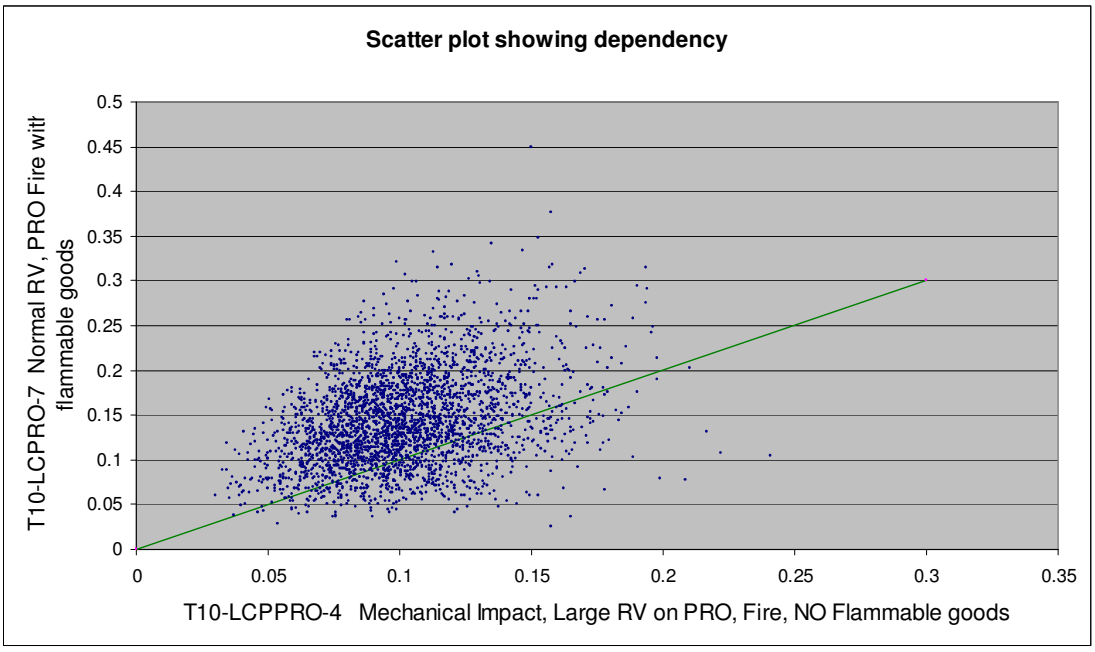


Fig. 6.4 Scatter plot of the two rule sets T10-LCPPRO-4 and T10-LCPPRO-7

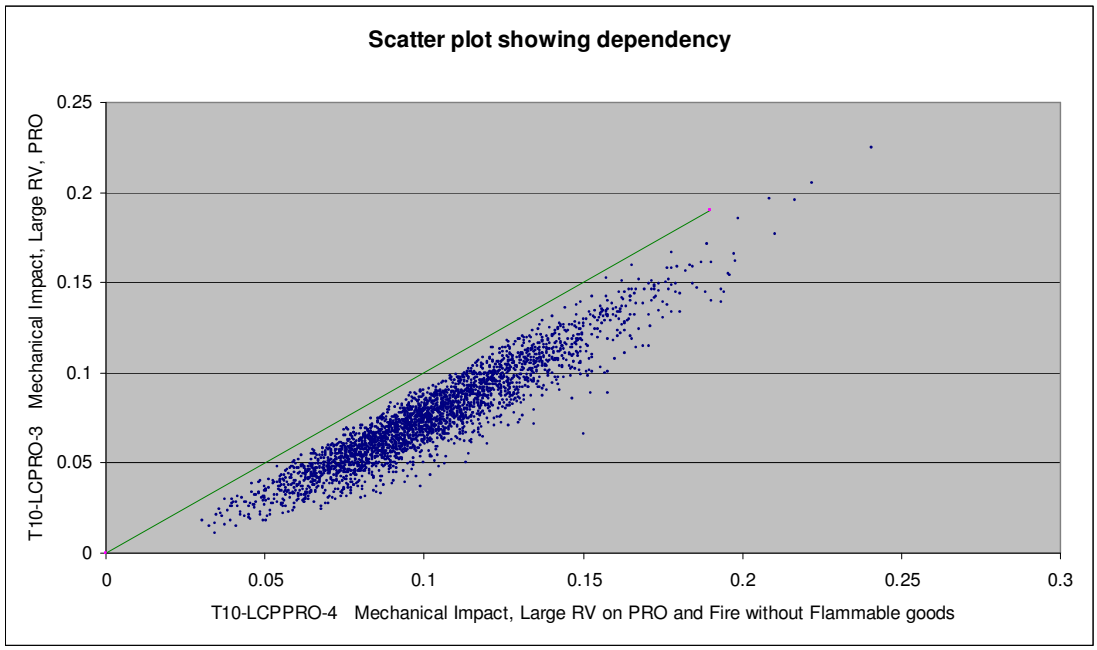


Fig. 6.5 Scatter plot of the two rule sets T10-LCPPRO-4 and T10-LCPPRO-3

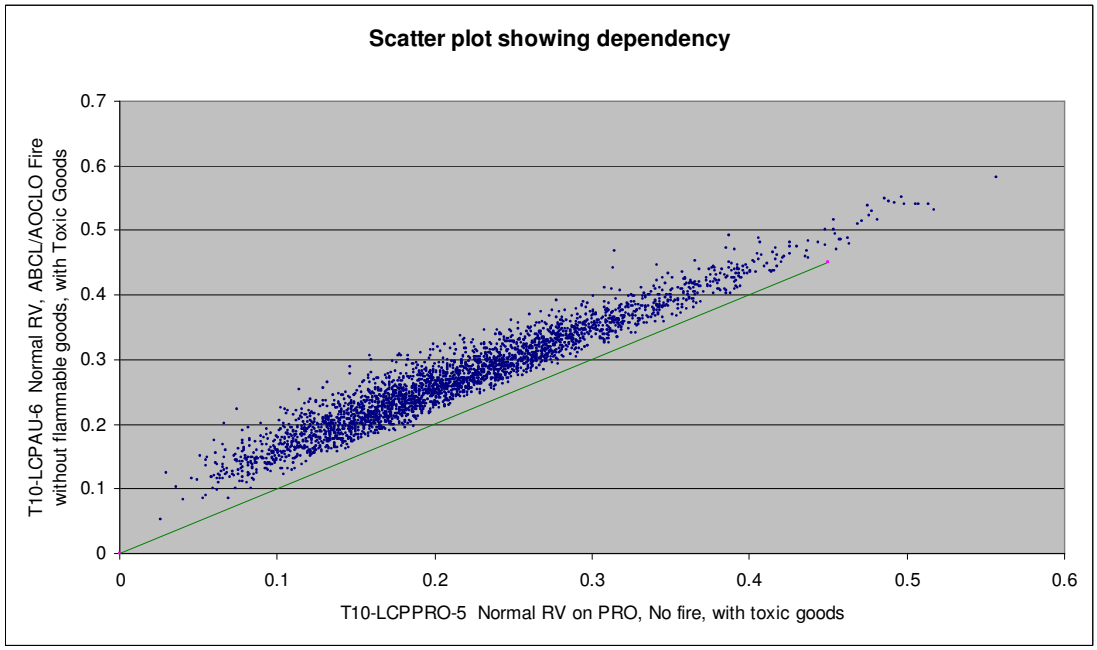


Fig. 6.6 Scatter plot of the two rule sets T10-LCPPRO-5 and T10-LCPAU-6

6.4.13 Uncertainty Assessment

We have built the covariance matrix of the precursors and the individual casualty probabilities of the rule sets related to HET10. We can then assess the uncertainty in the output of HET10 by the methods developed in Chapter 5.

At the first step, the mimic model is built on MS Excel. The mimic model is composed of 12,425 risk scenarios. With the mimic model, the analytical solutions of the mean and the standard deviation of the output are calculated as summarized in Table 6.33. The calculated analytical mean differentiates from the result from the SRM FT+ model by 0.0001 that means a relative error as small as $2.7e-5$. Monte-Carlo simulations are then conducted with 6000 sets of samples of the precursors and the rule sets. With the simulated outputs of HET10, the empirical mean and standard deviation are calculated as shown in Table 6.33. The relative errors compared with the calculated analytical solutions are 0.22% and 3.7% for the mean and the standard deviation respectively. The calculated analytical solutions and the simulations-based empirical solutions coincide very well and therefore verify each other. With the simulated outputs of HET10, the empirical density probability curve is drawn as shown in Fig. 6.7. The five vertical dash lines mark five percentiles as summarized in Table 6.34.

Table 6.33 Summary of the uncertainty assessment of HET10

Method	**Mean**	**Standard Deviation**
SRM FT+	3.7328	n/a
Analytical Solution	3.7327	0.7974
Monte-Carlo Simulations	3.7399	0.8271
Relative error	0.22%	3.7%

Table 6.34 Percentiles of the simulations of HET10

	5%	**25%**	**50%**	**75%**	**95%**
Percentile	2.344	3.181	3.734	4.285	5.060

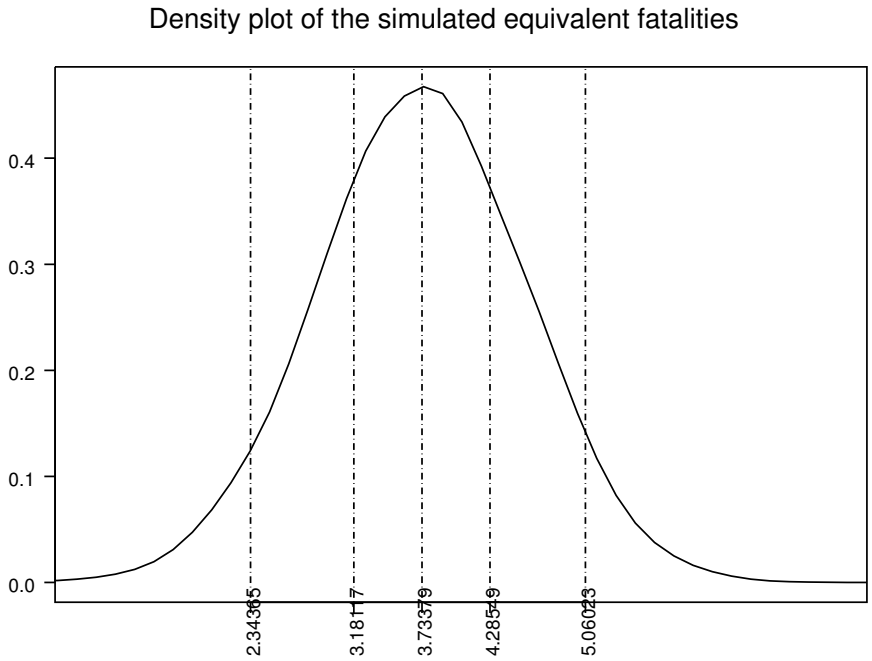


Fig. 6.7 Empirical density probability curve from the simulations of HET10

6.5 Case Study with HET12

6.5.1 Elicitation process for building the covariance matrix of the precursors

The hazardous event HET12 is defined as *the derailment of passenger trains*. It is the second largest contributor to the over all risk in RSSB-SRM [Dennis 2006]. There are 52 precursors related to HET12. To build up the covariance matrix of the precursors, we designed the elicitation process as shown in Fig. 6.8. At the first step, the precursor families are identified by the experts. After that, we elicit the uncertainty factors of all the precursors. The experts are then asked to identify the invariant factors for each precursor family. At the fourth step, the correlations regarding the uncertainty factors are elicited with the qualitative assessment plus the benchmark as designed in Chapter 3. The elicited correlations are monitored during the process by the indicator parameter q_i as defined in Chapter 3. As a result, the correlation matrix can be derived for the precursors. At the last step, the variance of each family is derived through elicitation by the method developed in Chapter 3. The

standard deviation of the precursors can then be derived within the families and the covariance matrix can then be derived.

We will continue to discuss the elicitations in details. To make the elicitation work easy to follow, the following subsections are headed according to the structure of Fig. 6.8. At the beginning of each subsection, we also locate the subsection with reference to the blocks in Fig. 6.8.

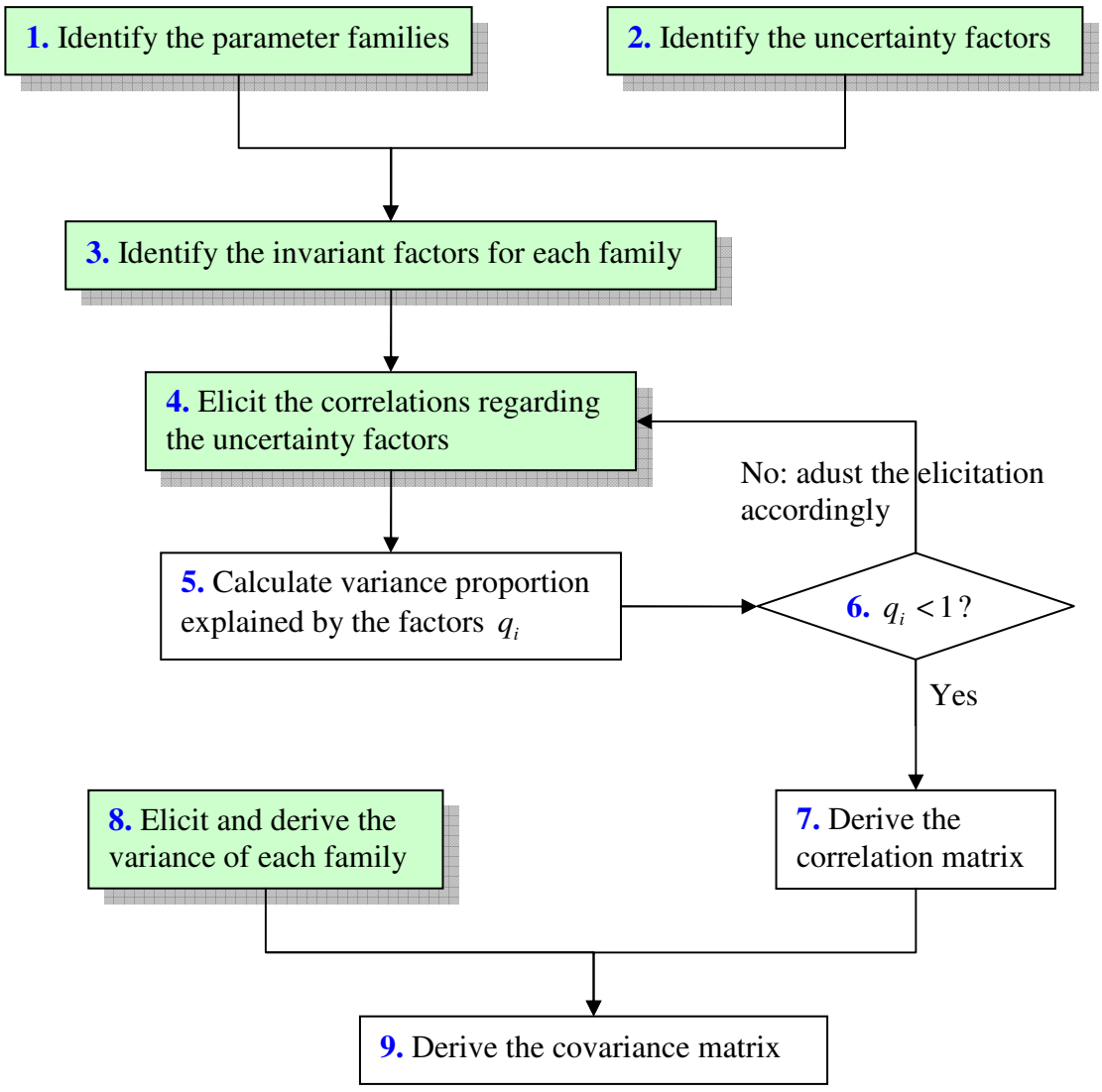


Fig. 6.8 Elicitation process for building up the covariance matrix for HET12 precursors

6.5.2 Precursor Families

This subsection is referred to Block 1 in Fig. 6.8. We will identify the precursor families.

The experts partition the 52 precursors related to HET12 by the definitions into 7 families that are related to: drivers' errors; environment issues; infrastructure defects; rolling stock defects; RV faults; staff errors; and track defects. The family codes, the associated precursors, and the descriptions of the precursors are summarized in the first three columns in Fig. 6.9.

6.5.3 Uncertainty Factors

This subsection is related to Block 2 in Fig. 6.8. We will elicit the uncertainty factors of the precursors.

For all the 52 precursors, 13 uncertainty factors are elicited as:

- Driver monitoring training;
- Extreme climate conditions;
- Network Rail performance;
- Signalling failures;
- Structural failure;
- Road vehicle driver error;
- Train maintenance;
- Vandalism;
- Public errors;
- Shunter training;
- Signaller error;
- Track maintenance;
- S&C maintenance.

6.5.4 Invariant Factors

This subsection is related to Block 3 in Fig. 6.8. We will elicit the invariant factors for each precursor family.

From the 13 uncertainty factors elicited in Subsection 6.5.4, the invariant factors for each family are elicited as summarized in Table 6.35.

Table 6.35 Invariant factors of the precursor families related to HET12

No.	Family Code	Invariant Factor
1.	Driver Error	Driver monitoring training
2.	Environment	Extreme climate conditions
3.	Infra	Network Rail
4.	Rolling Stock	Train maintenance
5.	RV	RV driver error
6.	Staff Error	Residual uncertainty
7.	Track Defect	Network Rail

6.5.5 Elicitations on the Uncertainty Factors

This subsection is related to Block 4 in Fig. 6.8. We will elicit the correlations between the precursors and the uncertainty factors.

The qualitative correlations between the precursors and the factors are elicited as summarized in Fig. 6.9. All these correlations are positive.

			Factor													
Family	Precusror	Description	Driver monitoring training	Extreme climate conditions	Network Rail	Signalling failures	Structural failure	Road driver error	Train maint.	Vandal-ism	Public error	Shunter training	Signaller error	Track maint.	S&C maint.
Driver Error	POSL----PH	Overspeeding leading to PT derailment	VS												
Driver Error	PSNT----PF	Severe braking/snatch leading to PT derailment	S						W						
Driver Error	PSPD---PH	SPAD at S&C leading to PT derailment	S												
Driver Error	UTRN----UE	Running into train derailed while in depots/sidings leading to train derailment	S												
Environment	BTRE----UE	Running into trees leading to train derailment		S	M										
Environment	RLNS----UF	Running into landslip leading to train derailment		S	M										
Environment	RQAK---UE	Structural damage due to earthquake leading to train derailment		M			S								
Environment	RSLP----PF	Subsidence/ landslip under track leading to PT derailment		S	M										
Environment	WFLD----UE	Running into flooding leading to train derailment		S	W										
Environment	WSNO----UE	Running into snow/ice leading to train derailment		M	S										
Environment	WWIN----UE	High winds leading to train derailment		S	M										
Environment	RSCR----UE	Rail bridge collapse - scour leading to train derailment		S	M										
Infra	RBGD----UF	Running into to debris from overbridges leading to train derailment			W		W	S							
Infra	RBGS---UF	Rail bridge structural failure leading to train derailment			M		M								
Infra	RBLD----UF	Running into debris from lineside structures/buildings leading to train derailment			M		M								
infra	RDRN----UF	Drainage culvert/pipework collapse leading to train derailment			M		M								
Infra	ROHL----UF	Running into debris from OHLE structures leading to train derailment			S		M								
Infra	RSIG----UF	Running into debris from signalling gantries leading to train derailment			S	M									

Fig. 6.9.a Elicited uncertainty factors of the precursors related to HET12, Part 1

			Factor													
Family	Precusror	Description	Driver monitoring training	Extreme climate conditions	Network Rail	Signalling failures	Structural failure	Road driver error	Train maint.	Vandal-ism	Public error	Shunter training	Signaller error	Track maint.	S&C maint.
Infra	RTUNWALLU	Running into debris in the tunnel leading to train			M		M								
Infra	RWAL----UF	Running into debris from retaining walls leading to train derailment			M		M								
Infra	SWRG----PF	Wrongside signal failure at S&C leading to PT derailment			M	M									
Rolling Stock	PAXB----PF	Seized axle box bearing leading to PT derailment							S						
Rolling Stock	PAXL----PF	Axle failure leading to PT derailment							S						
Rolling Stock	PBUF----PF	Buffer locking leading to PT derailment							M						
Rolling Stock	PCUP----PF	Coupling failure leading to PT derailment							M						
Rolling Stock	PSUS---PF	Suspension system/bogie failures leading to PT derailment							S						
Rolling Stock	PTRA----PF	Running into objects fallen from trains leading to PT derailment							M						
Rolling Stock	PWHF----PF	Wheel flats or wheel/tyre wear beyond limits leading to PT derailment							ES						
Rolling Stock	PWHL----PF	Wheel failure leading to PT Derailment							S						
RV	RBSHCOL-UE	Rail bridge collapse - bridge bashing leading to train derailment						S							
RV	RBSH----UE	Bridge bashing leading to bridge displacement (not collapse) and train derailment						S							
RV	VBGV----UE	Running into vehicles fallen from overbridge leading to train derailment						M		M					
RV	VBND----UE	Running into vehicles through boundary fence leading to train derailment						M		M					
Staff Error	BBLD----UE	Running into objects from building site leading to train derailment			VW					M	M				
Staff Error	MMAT----UE	Running into Engineers materials left foul leading to train derailment			W					W				S	

Fig. 6.9.b Elicited uncertainty factors of the precursors related to HET12, Part 2

			Factor												
Family	Precusror	Description	Driver monitoring training	Extreme climate conditions	Network Rail	Signalling failures	Structural failure	Road driver error	Train maint.	Vandal-ism	Public error	Shunter training	Signaller error	Track maint.	S&C maint.
Staff Error	PCATDR--PH	Cat D SPAD or runaway leading to PT derailment	M						M	M					
Staff Error	PCRW----PH	Other driver/train crew error at S&C leading to PT derailment	VS												
Staff Error	PSHN----PH	Shunter errors leading to PT derailment										VS			
Staff Error	TTKD----UE	Track damage from other undetected derailment leading to train derailment			M									S	
Staff Error	VMVE---UE	Running into maintenance vehicles leading to train derailment			M									S	
Staff Error	XPRD----PH	Track maintenance staff errors leading to PT derailment			W									S	
Staff Error	XSCO----PH	Incorrect scotch and clip of points leading to PT derailment			VW										VS
Staff Error	XSGM----PH	Signaller/ crossing keeper error leading to PT derailment			M								M		
Track Defect	TBCK----PF	Buckled rail leading to PT derailment		M	S									VW	
Track Defect	TBKR----PF	Broken rail leading to PT derailment			M									M	
Track Defect	TFSH----PF	Broken fishplate leading to PT derailment			M									M	
Track Defect	TSPG---PF	Gauge spread (assumed always slow speed) leading to PT derailment			M									M	
Track Defect	TTUNBKR-PF	Broken rail in tunnel leading to PT derailment			M									M	
Track Defect	TTWS----PF	Track twist leading to PT derailment			M									M	
Track Defect	XDSC----PF	Defective S&C leading to PT derailment			M										M
Track Defect	XPOS----PF	Points in the wrong position and not detected leading to PT derailment			M	M									VW
Track Defect	XSCM----PF	Movement of points under train (equipment faults) leading to PT derailment			M	VW									M

Fig. 6.9.c Elicited uncertainty factors of the precursors related to HET12, Part 3

6.5.6 Correlation Matrix

This subsection is related to Block 7 in Fig. 6.8. We will derive the correlation matrix of all the 52 precursors based on the elicitations regarding the uncertainty factors.

By the benchmarks given on Table 6.1, the qualitative correlations are mapped to numbers. The correlation matrix can then be built for the as shown in Fig. 6.10. For presentation, the correlation matrix is split into two 4 blocks as:

$$r(\mathrm{HET12}) = \left[\begin{array}{c|c} r_{1,1} & r_{1,2} \\ \hline r_{2,1} & r_{2,2} \end{array}\right].$$

The block $r_{1,1}$ represents the correlation matrix of the first group of 26 precursors as shown in Fig. 6.10.a. The block $r_{2,2}$ represents the correlation matrix of the second group of 26 precursors as shown in Fig. 6.10.b. The block $r_{1,2}$ represents the correlation matrix of between the two groups as shown in Fig. 6.10.c. The block $r_{2,1}$ is the transposition of $r_{1,2}$.

	POSL---PH	PSNT---PF	PSPD---PH	UTRN---UE	BTRE---UE	RLNS---UF	RQAK---UE	RSLP---PF	WFLD---UE	WSNO---UE	WWIN---UE	RBGD---UF	RBGS---UF	RBLD---UF	RDRN---UF	ROHL---UF	RSCR---UE	RSIG---UF	RTUNWALLUF	RWAL---UF	SWRG---PF	PAXB---PF	PAXL---PF	PBUF---PF	PCUP---PF	PSUS---PF
POSL----PH	1.00	0.73	0.73	0.73	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PSNT----PF	0.73	1.00	0.60	0.60	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.26	0.26	0.19	0.19	0.26
PSPD----PH	0.73	0.60	1.00	0.60	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
UTRN----UE	0.73	0.60	0.60	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BTRE---UE	0.00	0.00	0.00	0.00	1.00	0.92	0.44	0.92	0.79	0.87	0.92	0.92	0.19	0.32	0.32	0.32	0.44	0.44	0.32	0.32	0.32	0.00	0.00	0.00	0.00	0.00
RLNS----UF	0.00	0.00	0.00	0.00	0.92	1.00	0.44	0.92	0.79	0.87	0.92	0.92	0.19	0.32	0.32	0.32	0.44	0.44	0.32	0.32	0.32	0.00	0.00	0.00	0.00	0.00
RQAK----UE	0.00	0.00	0.00	0.00	0.44	0.44	1.00	0.44	0.44	0.32	0.44	0.44	0.26	0.44	0.44	0.44	0.44	0.00	0.44	0.44	0.00	0.00	0.00	0.00	0.00	0.00
RSLP----PF	0.00	0.00	0.00	0.00	0.92	0.92	0.44	1.00	0.79	0.87	0.92	0.92	0.19	0.32	0.32	0.32	0.44	0.44	0.32	0.32	0.32	0.00	0.00	0.00	0.00	0.00
WFLD----UE	0.00	0.00	0.00	0.00	0.79	0.79	0.44	0.79	1.00	0.69	0.79	0.79	0.11	0.19	0.19	0.19	0.26	0.26	0.19	0.19	0.19	0.00	0.00	0.00	0.00	0.00
WSNO----UE	0.00	0.00	0.00	0.00	0.87	0.87	0.32	0.87	0.69	1.00	0.87	0.87	0.26	0.44	0.44	0.44	0.60	0.60	0.44	0.44	0.44	0.00	0.00	0.00	0.00	0.00
WWIN----UE	0.00	0.00	0.00	0.00	0.92	0.92	0.44	0.92	0.79	0.87	1.00	0.92	0.19	0.32	0.32	0.32	0.44	0.44	0.32	0.32	0.32	0.00	0.00	0.00	0.00	0.00
RBGD----UF	0.00	0.00	0.00	0.00	0.92	0.92	0.44	0.92	0.79	0.87	0.92	1.00	0.19	0.32	0.32	0.32	0.44	0.44	0.32	0.32	0.32	0.00	0.00	0.00	0.00	0.00
RBGS----UF	0.00	0.00	0.00	0.00	0.19	0.19	0.26	0.19	0.11	0.26	0.19	0.19	1.00	0.37	0.37	0.37	0.44	0.26	0.37	0.37	0.19	0.00	0.00	0.00	0.00	0.00
RBLD----UF	0.00	0.00	0.00	0.00	0.32	0.32	0.44	0.32	0.19	0.44	0.32	0.32	0.37	1.00	0.63	0.63	0.75	0.44	0.63	0.63	0.32	0.00	0.00	0.00	0.00	0.00
RDRN----UF	0.00	0.00	0.00	0.00	0.32	0.32	0.44	0.32	0.19	0.44	0.32	0.32	0.37	0.63	1.00	0.63	0.75	0.44	0.63	0.63	0.32	0.00	0.00	0.00	0.00	0.00
ROHL----UF	0.00	0.00	0.00	0.00	0.32	0.32	0.44	0.32	0.19	0.44	0.32	0.32	0.37	0.63	0.63	1.00	0.75	0.44	0.63	0.63	0.32	0.00	0.00	0.00	0.00	0.00
RSCR----UE	0.00	0.00	0.00	0.00	0.44	0.44	0.44	0.44	0.26	0.60	0.44	0.44	0.44	0.75	0.75	0.75	1.00	0.60	0.75	0.75	0.44	0.00	0.00	0.00	0.00	0.00
RSIG----UF	0.00	0.00	0.00	0.00	0.44	0.44	0.00	0.44	0.26	0.60	0.44	0.44	0.26	0.44	0.44	0.44	0.60	1.00	0.44	0.44	0.75	0.00	0.00	0.00	0.00	0.00
RTUNWALLUF	0.00	0.00	0.00	0.00	0.32	0.32	0.44	0.32	0.19	0.44	0.32	0.32	0.37	0.63	0.63	0.63	0.75	0.44	1.00	0.63	0.32	0.00	0.00	0.00	0.00	0.00
RWAL----UF	0.00	0.00	0.00	0.00	0.32	0.32	0.44	0.32	0.19	0.44	0.32	0.32	0.37	0.63	0.63	0.63	0.75	0.44	0.63	1.00	0.32	0.00	0.00	0.00	0.00	0.00
SWRG----PF	0.00	0.00	0.00	0.00	0.32	0.32	0.00	0.32	0.19	0.44	0.32	0.32	0.19	0.32	0.32	0.32	0.44	0.75	0.32	0.32	1.00	0.00	0.00	0.00	0.00	0.00
PAXB----PF	0.00	0.26	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.60	0.44	0.44	0.60
PAXL----PF	0.00	0.26	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.60	1.00	0.44	0.44	0.60
PBUF----PF	0.00	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.44	0.44	1.00	0.32	0.44
PCUP----PF	0.00	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.44	0.44	0.32	1.00	0.44
PSUS----PF	0.00	0.26	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.60	0.60	0.44	0.44	1.00

Fig. 6.10.a Correlation matrix of the precursors related to HET12: Part 1

	PTRA---PF	PWHF---PF	PWHL---PF	RBSHCOL-UE	RBSH---UE	VBGV---UE	VBND---UE	BBLD---UE	MMAT---UE	PCATDR--PH	PCRW---PH	PSHN---PH	TTKD---UE	VMVE---UE	XPRD---PH	XSCO---PH	XSGM---PH	TBCK---PF	TBKR---PF	TFSH---PF	TSPG---PF	TTUNBKR-PF	TTWS---PF	XDSC---PF	XPOS---PF	XSCM---PF
PTRA---PF	1.00	0.56	0.44	0.00	0.00	0.00	0.00	0.00	0.00	0.32	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PWHF---PF	0.56	1.00	0.78	0.00	0.00	0.00	0.00	0.00	0.00	0.56	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PWHL---PF	0.44	0.78	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.44	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RBSHCOL-UE	0.00	0.00	0.00	1.00	0.60	0.44	0.44	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RBSH---UE	0.00	0.00	0.00	0.60	1.00	0.44	0.44	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
VBGV---UE	0.00	0.00	0.00	0.44	0.44	1.00	0.63	0.32	0.19	0.32	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
VBND---UE	0.00	0.00	0.00	0.44	0.44	0.63	1.00	0.32	0.19	0.32	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BBLD---UE	0.00	0.00	0.00	0.00	0.00	0.32	0.32	1.00	0.49	0.44	0.20	0.20	0.27	0.27	0.37	0.20	0.45	0.14	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
MMAT---UE	0.00	0.00	0.00	0.00	0.00	0.19	0.19	0.49	1.00	0.28	0.15	0.15	0.91	0.91	0.94	0.18	0.44	0.40	0.62	0.62	0.62	0.62	0.62	0.19	0.19	0.19
PCATDR--PH	0.32	0.56	0.44	0.00	0.00	0.32	0.32	0.44	0.28	1.00	0.60	0.08	0.06	0.06	0.12	0.06	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PCRW---PH	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.15	0.60	1.00	0.12	0.10	0.10	0.18	0.10	0.21	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PSHN---PH	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.15	0.08	0.12	1.00	0.10	0.10	0.18	0.10	0.21	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TTKD---UE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.27	0.91	0.06	0.10	0.10	1.00	1.00	0.94	0.18	0.49	0.58	0.75	0.75	0.75	0.75	0.75	0.32	0.32	0.32
VMVE---UE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.27	0.91	0.06	0.10	0.10	1.00	1.00	0.94	0.18	0.49	0.58	0.75	0.75	0.75	0.75	0.75	0.32	0.32	0.32
XPRD---PH	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.37	0.94	0.12	0.18	0.18	0.94	0.94	1.00	0.22	0.51	0.40	0.62	0.62	0.62	0.62	0.62	0.19	0.19	0.19
XSCO---PH	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.18	0.06	0.10	0.10	0.18	0.18	0.22	1.00	0.28	0.14	0.10	0.10	0.10	0.10	0.10	0.63	0.27	0.63
XSGM---PH	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.45	0.44	0.13	0.21	0.21	0.49	0.49	0.51	0.28	1.00	0.44	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32
TBCK---PF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.14	0.40	0.00	0.00	0.00	0.58	0.58	0.40	0.14	0.44	1.00	0.54	0.54	0.54	0.54	0.54	0.44	0.44	0.44
TBKR---PF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.62	0.00	0.00	0.00	0.75	0.75	0.62	0.10	0.32	0.54	1.00	0.63	0.63	0.63	0.63	0.32	0.32	0.32
TFSH---PF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.62	0.00	0.00	0.00	0.75	0.75	0.62	0.10	0.32	0.54	0.63	1.00	0.63	0.63	0.63	0.32	0.32	0.32
TSPG---PF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.62	0.00	0.00	0.00	0.75	0.75	0.62	0.10	0.32	0.54	0.63	0.63	1.00	0.63	0.63	0.32	0.32	0.32
TTUNBKR-PF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.62	0.00	0.00	0.00	0.75	0.75	0.62	0.10	0.32	0.54	0.63	0.63	0.63	1.00	0.63	0.32	0.32	0.32
TTWS---PF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.62	0.00	0.00	0.00	0.75	0.75	0.62	0.10	0.32	0.54	0.63	0.63	0.63	0.63	1.00	0.32	0.32	0.32
XDSC---PF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.19	0.00	0.00	0.00	0.32	0.32	0.19	0.63	0.32	0.44	0.32	0.32	0.32	0.32	0.32	1.00	0.42	0.63
XPOS---PF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.19	0.00	0.00	0.00	0.32	0.32	0.19	0.27	0.32	0.44	0.32	0.32	0.32	0.32	0.32	0.42	1.00	0.52
XSCM---PF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.19	0.00	0.00	0.00	0.32	0.32	0.19	0.63	0.32	0.44	0.32	0.32	0.32	0.32	0.32	0.63	0.52	1.00

Fig. 6.10.b Correlation matrix of the precursors related to HET12: Part 2

	PTRA---PF	PWHF---PF	PWHL---PF	RBSHCOL-UE	RBSH---UE	VBGV---UE	VBND---UE	BBLD---UE	MMAT---UE	PCATDR-PH	PCRW---PH	PSHN---PH	TTKD---UE	VMVE---UE	XPRD---PH	XSCO---PH	XSGM---PH	TBCK---PF	TBKR---PF	TFSH---PF	TSPG---PF	TTUNBKR-PF	TTWS---PF	XDSC---PF	XPOS---PF	XSCM---PF
POSL----PH	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.53	0.88	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PSNT----PF	0.19	0.33	0.26	0.00	0.00	0.00	0.00	0.00	0.00	0.62	0.73	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PSPD----PH	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.44	0.73	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
UTRN----UE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.44	0.73	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BTRE----UE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.19	0.00	0.00	0.00	0.32	0.32	0.19	0.10	0.32	0.87	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32
RLNS----UF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.19	0.00	0.00	0.00	0.32	0.32	0.19	0.10	0.32	0.87	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32
RQAK----UE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.32	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RSLP----PF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.19	0.00	0.00	0.00	0.32	0.32	0.19	0.10	0.32	0.87	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32
WFLD----UE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.11	0.00	0.00	0.00	0.19	0.19	0.11	0.06	0.19	0.69	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19
WSNO----UE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.14	0.26	0.00	0.00	0.00	0.44	0.44	0.26	0.14	0.44	0.92	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44
WWIN----UE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.19	0.00	0.00	0.00	0.32	0.32	0.19	0.10	0.32	0.87	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32
RBGD----UF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.19	0.00	0.00	0.00	0.32	0.32	0.19	0.10	0.32	0.87	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32
RBGS----UF	0.00	0.00	0.00	0.60	0.60	0.44	0.44	0.06	0.11	0.00	0.00	0.00	0.19	0.19	0.11	0.06	0.19	0.26	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19
RBLD----UF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.19	0.00	0.00	0.00	0.32	0.32	0.19	0.10	0.32	0.44	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32
RDRN----UF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.19	0.00	0.00	0.00	0.32	0.32	0.19	0.10	0.32	0.44	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32
ROHL----UF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.19	0.00	0.00	0.00	0.32	0.32	0.19	0.10	0.32	0.44	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32
RSCR----UE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.14	0.26	0.00	0.00	0.00	0.44	0.44	0.26	0.14	0.44	0.60	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44
RSIG----UF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.14	0.26	0.00	0.00	0.00	0.44	0.44	0.26	0.14	0.44	0.60	0.44	0.44	0.44	0.44	0.44	0.75	0.54
RTUNWALLUF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.19	0.00	0.00	0.00	0.32	0.32	0.19	0.10	0.32	0.44	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32
RWAL----UF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.19	0.00	0.00	0.00	0.32	0.32	0.19	0.10	0.32	0.44	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32
SWRG----PF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.19	0.00	0.00	0.00	0.32	0.32	0.19	0.10	0.32	0.44	0.32	0.32	0.32	0.32	0.32	0.32	0.63	0.42
PAXB----PF	0.44	0.78	0.60	0.00	0.00	0.00	0.00	0.00	0.00	0.44	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PAXL----PF	0.44	0.78	0.60	0.00	0.00	0.00	0.00	0.00	0.00	0.44	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PBUF----PF	0.32	0.56	0.44	0.00	0.00	0.00	0.00	0.00	0.00	0.32	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PCUP----PF	0.32	0.56	0.44	0.00	0.00	0.00	0.00	0.00	0.00	0.32	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PSUS----PF	0.44	0.78	0.60	0.00	0.00	0.00	0.00	0.00	0.00	0.44	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Fig. 6.10.c Correlation matrix of the precursors related to HET12: Part 3

6.5.7 Variance of Each Family

This subsection is referred to Block 8 in Fig. 6.8. We will elicit the variance of each of the seven precursor families as elicited in Subsection 6.5.2.

By the method developed in Chapter 3, the 95 percentile of the waiting time is elicited for the seven families as summarized in Table 6.36. The standard deviations are then derived as summarized in Table 6.36.

Table 6.36 Elicited 95 percentile of the waiting time, i.e. $t_{0.95}$, and the derived standard deviation of the occurrence rate of the families related to HET12

Family Code	Minimum $t_{0.95}$ (Months)	Maximum $t_{0.95}$ (Months)	Assessment of $t_{0.95}$ (Months)	Std σ_λ
Driver Error	55.8	63.5	59.5	4.69E-10
Environment	23.0	26.0	24.5	1.14E-09
Infra	10.0	11.3	10.6	2.62E-09
Rolling Stock	37.6	42.8	40	6.95E-10
RV	77.0	87.6	82	3.40E-10
Staff Error	24.0	27.3	25.5	1.09E-09
Track Defect	8.1	9.2	8.7	3.22E-09

6.5.8 Covariance Matrix

This subsection is related to Block 9 in Fig. 6.8. We will derive the standard deviations of the precursors and consequently derive the covariance matrix of all the 52 precursors.

Based one the family variance as summarized in Table 6.36 and the elicitations regarding the uncertainty factors, the standard deviations of the precursors are derived within the associated families as summarized in Table 6.37.

With the standard deviations summarized in Table 6.37 and the correlation matrix presented in Fig. 6.10, the covariance matrix of the 52 precursors can be built straightforwardly and therefore is not presented here.

Table 6.37 Standard deviation of the 52 precursors related to HET12

Family	Precursors	Mean	STD
Driver Error	POSL----PH	3.060E-10	5.480E-11
Driver Error	PSNT----PF	1.801E-10	3.906E-11
Driver Error	PSPD----PH	1.836E-09	3.982E-10
Driver Error	UTRN----UE	1.836E-11	3.982E-12
Environment	BTRE----UE	1.224E-09	2.472E-10
Environment	RLNS----UF	3.672E-09	7.417E-10
Environment	RQAK----UE	3.602E-13	1.002E-13
Environment	RSLP----PF	5.771E-10	1.166E-10
Environment	WFLD----UE	1.801E-13	3.638E-14
Environment	WSNO----UE	2.295E-10	6.385E-11
Environment	WWIN----UE	1.836E-12	3.708E-13
Infra	RBGD----UF	2.295E-10	3.502E-11
Infra	RBGS----UF	7.205E-12	2.582E-12
Infra	RBLD----UF	2.295E-10	4.824E-11
infra	RDRN----UF	1.836E-10	3.859E-11
Infra	ROHL----UF	1.836E-12	3.859E-13
Infra	RSCR----UE	9.006E-11	1.374E-11
Infra	RSIG----UF	1.836E-13	2.802E-14
Infra	RTUNWALLUF	1.191E-08	2.504E-09
Infra	RWAL----UF	3.672E-11	7.718E-12
Infra	SWRG----PF	3.846E-10	8.084E-11
Rolling Stock	PAXB----PF	4.246E-10	8.616E-11
Rolling Stock	PAXL----PF	2.123E-10	4.308E-11
Rolling Stock	PBUF----PF	2.123E-10	5.934E-11
Rolling Stock	PCUP----PF	3.602E-11	1.007E-11
Rolling Stock	PSUS----PF	2.123E-10	4.308E-11

Family	Precursors	Mean	STD
Rolling Stock	PTRA----PF	1.801E-09	5.034E-10
Rolling Stock	PWHF----PF	1.201E-10	1.890E-11
Rolling Stock	PWHL----PF	4.503E-10	9.137E-11
RV	RBSHCOL-UE	2.161E-11	3.621E-12
RV	RBSH----UE	2.202E-10	3.689E-11
RV	VBGV----UE	2.295E-10	5.296E-11
RV	VBND----UE	1.224E-09	2.824E-10
Staff Error	BBLD----UE	7.344E-12	1.748E-12
Staff Error	MMAT----UE	2.295E-10	7.432E-11
Staff Error	PCATDR--PH	2.295E-10	1.429E-10
Staff Error	PCRW----PH	5.771E-10	2.311E-10
Staff Error	PSHN----PH	2.295E-10	9.189E-11
Staff Error	TTKD----UE	7.205E-11	3.473E-11
Staff Error	VMVE----UE	1.836E-11	8.849E-12
Staff Error	XPRD----PH	5.771E-10	1.475E-10
Staff Error	XSCO----PH	4.328E-10	2.033E-10
Staff Error	XSGM----PH	3.060E-09	6.955E-10
Track Defect	TBCK----PF	1.154E-09	2.217E-10
Track Defect	TBKR----PF	1.154E-09	3.054E-10
Track Defect	TFSH----PF	5.662E-10	1.498E-10
Track Defect	TSPG----PF	3.060E-09	8.097E-10
Track Defect	TTUNBKR-PF	5.771E-09	1.527E-09
Track Defect	TTWS----PF	6.755E-10	1.787E-10
Track Defect	XDSC----PF	3.060E-09	8.097E-10
Track Defect	XPOS----PF	4.328E-10	1.145E-10
Track Defect	XSCM----PF	2.164E-10	5.726E-11

6.5.9 Uncertainty Models in the Rule Sets

6.5.9.1 Hazards

There are 123 rule sets associated with the consequences in HET12. The rule sets are summarized in Appendix B. Seven injury sources are elicited as: (1) train derailment; (2) carriage on its side; (3) structure collapsed onto the train; (4) secondary collision after derailment; (5) fire; (6) toxic goods; (7) train falling down the bridge. The injury sources related to HET12 are mainly affected by the locations on the track. Different locations of the track are associated with different derailment speeds and the different equipment structures. The locations are classified into: Bridge (BG), Open Fast (OF), Open Slow (OS), Single Track Tunnel (ST), Twin Track Tunnel (TT), and Station Fast (SF). The injury source of fire is also affected by whether there are extra flammable goods. Depending on the locations on the track and whether there are flammable goods, the hazards related to HET12 are defined as summarized in Table 6.38.

Table 6.38 Hazards related to HET12.

Hazard Types	Hazard specification	Weight
Falling down bridge	Falling down bridge	1
Derailment	On open track	0.333
	In tunnel	1
Carriage on its side after derailment	On open slow track	0.303
	On open fast AUTO track with the average speed of derailment assumed as 50mph	0.356
	On open fast track with the average speed of derailment assumed as 55mph.	0.541
	In twin track tunnel	1
Structure collapsed onto train following the collision	On open slow track	0.4
	On open fast track	1
Second collision after the derailment	Carriage hitting line side structure when train not on its side; on open slow track	0.010
	Carriage hitting line side structure when train not on its side; on open fast AUTO track with the average derailment speed assumed as 50mph	0.028
	Carriage hitting line side structure when train not on its side; on open fast track with the average derailment speed assumed as 55mph	0.046

Hazard Types	Hazard specification	Weight
	Carriage hitting line side structure when train on its side; on open slow track	0.020
	Carriage hitting line side structure when train on its side; on open fast track	0.133
	Carriage in contact with tunnel wall	0.181
	In collision with platform at a station	0.067
	In collision with another train; on open track	0.667
	In collision with another train; at a station	0.767
	In collision with another train; in twin track tunnel	1
Fire	Either in collision on line side structure or with carriage on its side; open slow track	0.016
	Either in collision on line side structure or with carriage on its side; open fast track	0.032
	The structure in collision collapsed onto the train on open slow track	0.016
	The structure in collision collapsed onto the train on open fast track	0.032
	Fire at a station	0.008
	In collision with a train without flammable goods on open track	0.040
	In collision with a train without flammable goods at a station	0.048
	With flammable goods released from the freight train in collision on open track	0.159
	With flammable goods from the freight train in collision at a station	0.254
	In tunnel	1
Toxic goods	Toxic goods release from the freight train in collision on open track	0.238
	Toxic goods release from the freight train in collision at a station	0.317
	Toxic goods release from the freight train in collision in twin track tunnel	1

6.5.9.2 Definitions of the Injury Atoms and the Rule Sets

For each hazard source, the injury atom is defined by the hazard with the worst hazard level. The weights of the hazards relative to the associated injury atoms are included in Table 6.38. We have assumed that the injury atoms follow Beta distributions. By the methods developed in Chapter 4, the definition parameters of the seven injury atoms are derived as summarized in Table 6.39.

On the injury atoms defined in Table 6.39, the 123 rule sets related to HET12 are then defined as summarized in Table 6.40.

Table 6.39 Definitions of the injury atoms related to HET12

Injury Source	Code	The Mean	99 percentile	α	β	Std
Derailment	DERTNL	0.0043	0.0086	7.95	1846	0.002
Carriage on side	CARONSTNL	0.2427	0.4855	5.02	15.7	0.092
Structure collapsed onto the train	STROnTNOS	0.25	0.5	4.93	14.8	0.095
Second collision	SecCollTNL	0.2143	0.4286	5.38	19.7	0.080
Fire	FireTNL	0.90	0.99	20.31	2.26	0.062
Toxic goods	ToxicTNL	0.90	0.99	20.31	2.26	0.062
Falling off a bridge	onBridge	0.1429	0.2857	6.27	37.6	0.052

Table 6.40 Definitions of the rule sets related to HET12 on the injury atoms.

Rule Sets	DER TNL	CARON STNL	STRO n TNOS	SecColl TNL	FireTNL	Toxic TNL	On Bridge
T12-BG-1	0.000	0.000	0.000	0.000	0.000	0.000	1.000
T12-OF-01	0.333	0.000	0.000	0.000	0.000	0.000	0.000
T12-OF-02	0.333	0.000	0.000	0.000	0.000	0.000	0.000
T12-OF-03	0.333	0.000	0.000	0.046	0.000	0.000	0.000
T12-OF-04	0.333	0.000	0.000	0.046	0.032	0.000	0.000
T12-OF-05	0.333	0.000	1.000	0.046	0.000	0.000	0.000
T12-OF-06	0.333	0.000	1.000	0.046	0.032	0.000	0.000
T12-OF-07	0.333	0.541	0.000	0.000	0.000	0.000	0.000
T12-OF-08	0.333	0.541	0.000	0.000	0.032	0.000	0.000
T12-OF-09	0.333	0.541	0.000	0.133	0.000	0.000	0.000
T12-OF-10	0.333	0.541	0.000	0.133	0.032	0.000	0.000
T12-OF-11	0.333	0.541	0.657	0.133	0.000	0.000	0.000
T12-OF-12	0.333	0.541	0.657	0.133	0.048	0.000	0.000
T12-OF-13	0.333	0.000	0.000	0.667	0.000	0.000	0.000
T12-OF-14	0.333	0.000	0.000	0.667	0.040	0.000	0.000
T12-OF-15	0.333	0.000	0.000	0.667	0.000	0.000	0.000
T12-OF-16	0.333	0.000	0.000	0.667	0.040	0.000	0.000
T12-OF-17	0.333	0.000	0.000	0.667	0.000	0.238	0.000
T12-OF-18	0.333	0.000	0.000	0.667	0.040	0.238	0.000
T12-OF-19	0.333	0.000	0.000	0.667	0.000	0.000	0.000
T12-OF-20	0.333	0.000	0.000	0.667	0.198	0.000	0.000
T12-OF-21	0.333	0.000	0.000	0.667	0.000	0.000	0.000
T12-OF-22	0.333	0.000	0.000	0.667	0.040	0.000	0.000

T12-OF-23	0.333	0.000	0.000	0.667	0.000	0.000	0.000
T12-OF-24	0.333	0.000	0.000	0.667	0.040	0.000	0.000
T12-OFAUTO-01	0.333	0.000	0.000	0.000	0.000	0.000	0.000
T12-OFAUTO-02	0.333	0.000	0.000	0.000	0.000	0.000	0.000
T12-OFAUTO-03	0.333	0.000	0.000	0.028	0.000	0.000	0.000
T12-OFAUTO-04	0.333	0.000	0.000	0.028	0.032	0.000	0.000
T12-OFAUTO-05	0.333	0.000	1.000	0.028	0.000	0.000	0.000
T12-OFAUTO-06	0.333	0.000	1.000	0.028	0.032	0.000	0.000
T12-OFAUTO-07	0.333	0.356	0.000	0.000	0.000	0.000	0.000
T12-OFAUTO-08	0.333	0.356	0.000	0.000	0.032	0.000	0.000
T12-OFAUTO-09	0.333	0.356	0.000	0.133	0.000	0.000	0.000
T12-OFAUTO-10	0.333	0.356	0.000	0.133	0.032	0.000	0.000
T12-OFAUTO-11	0.333	0.356	0.657	0.133	0.000	0.000	0.000
T12-OFAUTO-12	0.333	0.356	0.657	0.133	0.048	0.000	0.000
T12-OFAUTO-13	0.333	0.000	0.000	0.667	0.000	0.000	0.000
T12-OFAUTO-14	0.333	0.000	0.000	0.667	0.040	0.000	0.000
T12-OFAUTO-15	0.333	0.000	0.000	0.667	0.000	0.000	0.000
T12-OFAUTO-16	0.333	0.000	0.000	0.667	0.040	0.000	0.000
T12-OFAUTO-17	0.333	0.000	0.000	0.667	0.000	0.238	0.000
T12-OFAUTO-18	0.333	0.000	0.000	0.667	0.040	0.238	0.000
T12-OFAUTO-19	0.333	0.000	0.000	0.667	0.000	0.000	0.000
T12-OFAUTO-20	0.333	0.000	0.000	0.667	0.198	0.000	0.000
T12-OFAUTO-21	0.333	0.000	0.000	0.667	0.000	0.000	0.000
T12-OFAUTO-22	0.333	0.000	0.000	0.667	0.040	0.000	0.000
T12-OFAUTO-23	0.333	0.000	0.000	0.667	0.000	0.000	0.000
T12-OFAUTO-24	0.333	0.000	0.000	0.667	0.040	0.000	0.000
T12-OS-03	0.000	0.000	0.000	0.010	0.000	0.000	0.000
T12-OS-04	0.000	0.000	0.000	0.010	0.016	0.000	0.000
T12-OS-05	0.000	0.000	0.400	0.010	0.000	0.000	0.000
T12-OS-06	0.000	0.000	0.400	0.010	0.016	0.000	0.000
T12-OS-07	0.000	0.030	0.000	0.000	0.000	0.000	0.000
T12-OS-08	0.000	0.030	0.000	0.000	0.032	0.000	0.000
T12-OS-09	0.000	0.030	0.000	0.020	0.000	0.000	0.000
T12-OS-10	0.000	0.030	0.000	0.020	0.032	0.000	0.000
T12-OS-11	0.000	0.030	0.400	0.020	0.000	0.000	0.000
T12-OS-12	0.000	0.030	0.400	0.020	0.016	0.000	0.000
T12-OS-13	0.000	0.000	0.000	0.667	0.000	0.000	0.000
T12-OS-14	0.000	0.000	0.000	0.667	0.040	0.000	0.000
T12-OS-15	0.000	0.000	0.000	0.667	0.000	0.000	0.000
T12-OS-16	0.000	0.000	0.000	0.667	0.040	0.000	0.000
T12-OS-17	0.000	0.000	0.000	0.667	0.000	0.238	0.000
T12-OS-18	0.000	0.000	0.000	0.667	0.040	0.238	0.000
T12-OS-19	0.000	0.000	0.000	0.667	0.000	0.000	0.000
T12-OS-20	0.000	0.000	0.000	0.667	0.198	0.000	0.000
T12-OS-21	0.000	0.000	0.000	0.667	0.000	0.000	0.000
T12-OS-22	0.000	0.000	0.000	0.667	0.040	0.000	0.000
T12-OS-23	0.000	0.000	0.000	0.667	0.000	0.000	0.000

T12-OS-24	0.000	0.000	0.000	0.667	0.040	0.000	0.000
T12-ST-1	1.000	0.000	0.000	0.000	0.000	0.000	0.000
T12-ST-2	1.000	0.000	0.000	0.181	0.000	0.000	0.000
T12-ST-3	0.000	0.000	0.000	0.000	1.000	0.000	0.000
T12-TT-1	1.000	0.000	0.000	0.000	0.000	0.000	0.000
T12-TT-2	1.000	0.000	0.000	0.181	0.000	0.000	0.000
T12-TT-3	0.000	0.000	0.000	0.000	1.000	0.000	0.000
T12-TT-7	1.000	1.000	0.000	0.000	0.000	0.000	0.000
T12-TT-8	0.000	0.000	0.000	0.000	1.000	0.000	0.000
T12-TT-13	0.333	0.000	0.000	1.000	0.000	0.000	0.000
T12-TT-14	0.000	0.000	0.000	0.000	1.000	0.000	0.000
T12-TT-15	0.333	0.000	0.000	1.000	0.000	0.000	0.000
T12-TT-16	0.000	0.000	0.000	0.000	1.000	0.000	0.000
T12-TT-17	0.000	0.000	0.000	0.000	0.000	1.000	0.000
T12-TT-18	0.000	0.000	0.000	0.000	0.000	1.000	0.000
T12-TT-19	0.333	0.000	0.000	1.000	0.000	0.000	0.000
T12-TT-20	0.000	0.000	0.000	0.000	1.000	0.000	0.000
T12-TT-21	0.333	0.000	0.000	1.000	0.000	0.000	0.000
T12-TT-22	0.000	0.000	0.000	0.000	1.000	0.000	0.000
T12-TT-23	0.333	0.000	0.000	1.000	0.000	0.000	0.000
T12-TT-24	0.000	0.000	0.000	0.000	1.000	0.000	0.000
T14-SF-1	0.333	0.000	0.000	0.000	0.000	0.000	0.000
T14-SF-2	0.333	0.000	0.000	0.067	0.000	0.000	0.000
T14-SF-3	0.333	0.000	0.000	0.112	0.000	0.000	0.000
T14-SF-4	0.333	0.000	0.000	0.112	0.040	0.000	0.000
T14-SF-7	0.333	0.541	0.000	0.067	0.000	0.000	0.000
T14-SF-8	0.333	0.541	0.000	0.067	0.040	0.000	0.000
T14-SF-9	0.333	0.541	0.000	0.200	0.000	0.000	0.000
T14-SF-10	0.333	0.541	0.000	0.200	0.040	0.000	0.000
T14-SF-13	0.333	0.000	0.000	0.767	0.000	0.000	0.000
T14-SF-14	0.333	0.000	0.000	0.767	0.048	0.000	0.000
T14-SF-15	0.333	0.000	0.000	0.767	0.000	0.000	0.000
T14-SF-16	0.333	0.000	0.000	0.767	0.048	0.000	0.000
T14-SF-17	0.333	0.000	0.000	0.667	0.000	0.317	0.000
T14-SF-18	0.333	0.000	0.000	0.667	0.040	0.317	0.000
T14-SF-19	0.333	0.000	0.000	0.767	0.000	0.000	0.000
T14-SF-20	0.333	0.000	0.000	0.767	0.254	0.000	0.000
T14-SF-21	0.333	0.000	0.000	0.767	0.000	0.000	0.000
T14-SF-22	0.333	0.000	0.000	0.767	0.040	0.000	0.000
T14-SF-23	0.333	0.000	0.000	0.767	0.000	0.000	0.000
T14-SF-24	0.333	0.000	0.000	0.767	0.008	0.000	0.000
T10-LCPAU-2	0.000	0.000	0.000	0.000	0.032	0.000	0.000
T10-LCPAU-3	0.000	0.000	0.000	0.051	0.000	0.000	0.000
T10-LCPAU-4	0.000	0.000	0.000	0.051	0.032	0.000	0.000
T10-LCPAU-5	0.000	0.000	0.000	0.000	0.000	0.238	0.000
T10-LCPAU-6	0.000	0.000	0.000	0.000	0.079	0.238	0.000
T10-LCPAU-7	0.000	0.000	0.000	0.000	0.159	0.000	0.000

T10-LCPRO-2	0.000	0.000	0.000	0.000	0.032	0.000	0.000
T10-LCPRO-3	0.000	0.000	0.000	0.352	0.000	0.000	0.000
T10-LCPRO-4	0.000	0.000	0.000	0.352	0.032	0.000	0.000
T10-LCPRO-5	0.000	0.000	0.000	0.000	0.000	0.238	0.000
T10-LCPRO-6	0.000	0.000	0.000	0.000	0.079	0.238	0.000
T10-LCPRO-7	0.000	0.000	0.000	0.000	0.159	0.000	0.000

6.5.10 Uncertainty Assessment

We have built the covariance matrix for the precursors and the uncertainty model of the rule sets. We can then assess the uncertainty in the output of HET12 by the methods developed in Chapter 5.

At the first step, the mimic model is built on MS Excel. The mimic model is composed of 28132 risk scenarios. The analytical solutions of the mean and the standard deviation are calculated as summarized in Table 6.41. The calculated analytical mean differentiates from the result from the FT+ model by 0.008 that means a relative error as small as $2.7e-3$. Monte-Carlo simulations are then conducted with 6000 sets of samples of the precursors and the rule sets. With the simulated outputs of HET12, the empirical mean and standard deviation are calculated as shown in Table 6.41, of which the relative errors compared with the calculated analytical solutions are 0.31% and 0.56% respectively. The calculated analytical solutions and the simulations-based empirical solutions coincide very well and therefore verify each other. With the simulated outputs, the empirical density probability curve of HET12 is drawn as shown in Fig. 6.11. The five vertical dash lines mark the five percentiles as summarized in Table 6.42.

Table 6.41 Summary of the uncertainty assessment of HET12

Method	Mean	Standard Deviation
SRM FT+	3.136422777	n/a
Analytical Solution	3.127937928	0.623461051
Monte-Carlo Simulations	3.11814552	0.619988727
Relative error	0.31%	0.56%

Table 6.42 Percentiles of the simulations of HET12

	5%	25%	50%	75%	95%
Percentile	2.19	2.68	3.06	3.51	4.21

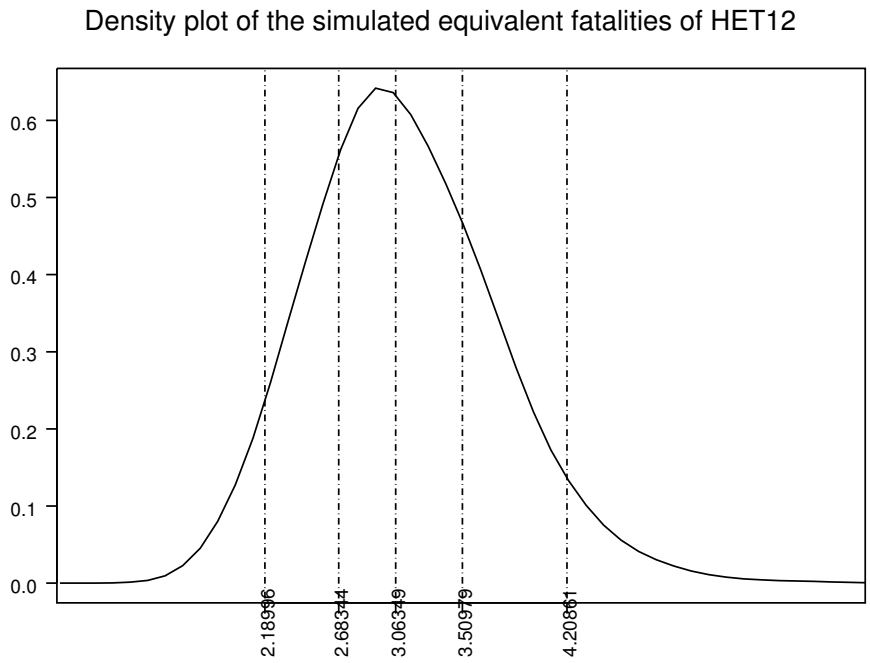


Fig. 6.11 Empirical density probability curve from the simulations of HET12

6.6 Conclusion

In this chapter, we have conducted two case studies on the RSSB-SRM hazardous event HET10 and HET12 respectively. We build the covariance matrices for the precursors by the procedure developed in Chapter 3. The case studies show that the experts are satisfied to use uncertainty factors as a structure to define the correlations among a set of input parameters. The experts feel also confident to make the qualitative assessment of the correlations regarding the uncertainty factors. In conjunction with the benchmark technique as used in *Crystal Ball* [ORACLE 2008], qualitative assessment proves an efficient way for eliciting the correlations regarding a large number of uncertainty factors. By the procedure developed in Chapter 3, the assessments of the correlations regarding the uncertainty factors are monitored

during the process. The case studies show that such a monitoring scheme can help the experts to adjust their assessment. With the monitoring scheme, the outcome correlation matrix can be guaranteed to be positive semi-definite. For each family, the percentile of the waiting time is elicited for deriving the variance. The standard deviations of the precursors are then derived within the associated family. The case studies demonstrate that the experts are confident to make the assessment on the waiting time percentile of a precursor family rather than the individual precursors.

The uncertainty in the rule sets is modelled on the injury atoms by the methods developed in Chapter 4 for the two case studies. The case studies demonstrate that the experts are confident in identifying the hazard types and structuring the rule sets into the injury atoms. The case studies demonstrate also that modelling on the injury atoms is an efficient way to model the uncertainty in the rule sets including the dependence and the order relationship among the rule sets.

Based on the uncertainty models in the precursors and the rule sets built previously, the uncertainty of the output is assessed by the methods developed in Chapter 6 for HET10 and HET12 respectively. The case studies demonstrate that the Excel mimic model is a reliable and efficient way to representing the Fault-tree and Event-tree models. With the case studies, the analytical solutions of the variance match the results from the simulations very well. It shows that the analytical solution is a robust way even for a large mimic model of tens of thousands of risk scenarios. The case studies also show that the software tools are correct and reliable.

As the RSSB-SRM HET10 and HET12 represent typical Fault-tree and Event-tree models, the case studies demonstrate that the procedure and the methods developed in Chapters 3, 4 and 5 are efficient for conducting the subjective uncertainty analysis of Fault-tree and Event-tree models.

Chapter 7

Conclusion

7.1 Review of the Aims of the Research

7.1.1 Overview of the Context of This Research

The goal of this research is to assess the subjective uncertainty in the output of a PRA model composed of fault trees and event trees. As shown in Fig. 1.8, a PRA model is usually built up on three layers. In the inner layer are the fault trees and event trees that are generally built up with computer software tools such as Isograph FT+. Therefore these fault trees and event trees are represented as computers codes. The fault trees and event trees are then parametralized; the database of the input parameters forms the parameter layer around the inner layer of the fault trees and event trees. In the context of risk analysis, the parameters are usually assessed through expert judgement elicitation. The experts usually make some assumptions from where to assess the parameters. The assumptions are narrated outside the parameter layer.

The uncertainty analysis in this research is conducted at the parameter layer. We cannot assess with certainty the value of each input parameters. Our uncertainty in the parameters is due to the lack of knowledge and therefore is categorized into epistemic uncertainty [Bedford and Cooke 2001]. For uncertainty analysis, we need firstly to model the uncertainty in the input parameters. In this research we focus on studying the uncertainty in the basic events and the consequences while we treat as constants the numbers of people exposed to risk scenarios and the probabilities of the outcomes of the escalation events. After building up the uncertainty in the input

parameters, we need to compute the uncertainty in the output, which is the uncertainty propagation of the uncertainty in the input parameters through the PRA model.

According to the above procedure, we identify three aims for this research as to be discussed in details.

7.1.2 Aim 1: A Procedure for Building the Covariance Matrix through Expert Judgement Elicitation

For a large group of input parameters, it is difficult to build the uncertainty model through expert judgment elicitation. In practice, it is usually split into two tasks: (1) building the marginal distributions for all the individual parameters; and (2) building the dependence among all the parameters [Kurowicka and Cooke 2006]. When the normal distribution is assumed as the marginal for each single parameter, the uncertainty in a group of input parameters can be completely defined by the means and the covariance matrix.

It is however a challenge to build up the covariance matrix for a large group of input parameters through expert judgement elicitation. Mainly there are two major issues. First, it requires too much elicitation workload from the experts. For n random variables, there are $n(n-1)/2$ correlation coefficients needed to elicit to build the correlation matrix. We need also to elicit the variance of each input parameter to build up the covariance matrix from the correlation matrix. Second, it is even more difficult to keep the outcome correlation matrix positive definite or positive semi-definite. Different methods have been developed to obtain a positive semi-definite correlation matrix by adjusting the outcome matrix from elicitation. For this purpose, however, arbitrary information has to be introduced [Ghosh and Henderson 2002; Ghosh and Henderson 2003; Kurowicka and Cooke 2006].

According to the above issues, we aim to develop a procedure for building the covariance matrix through expert judgement elicitation. The procedure should

require reasonable elicitation time from experts. The procedure should also be able to guarantee that the outcome covariance matrix is positive semi-definite.

7.1.3 Aim 2: Methods for Modelling the Uncertainty in the Consequences of an Event Tree

An event tree starts with an initiating event that has the potential to cause safety risk. A set of events follows the initiating event. These events can take different outcomes that affect the final consequences. Therefore they are called escalation events. A path from the initiating event through all the escalation events is called an accident sequence. The risk scenario is specified at the end of each accident sequence. Each individual exposed to the risk suffers a probability of getting injured at one of the four levels including: no injuries, minor injuries, major injuries and fatalities. We assume that all the individuals exposed to one risk scenario get injured independently. The consequences associated with a risk scenario are defined as the means of injuries at the four levels, i.e. the product of the number of people exposed to the risk scenario and the individual injury probabilities. Consequently the uncertainty in the consequences can be modelled on the uncertainty in the number of the people exposed to the risk scenario and the uncertainty in the individual injury probabilities. In this research, we don't study the uncertainty in the number of people exposed to the risk scenario as presented before. We focus on modelling the uncertainty in the individual injury probabilities.

The escalation events that affect the individual injury probabilities form a subset of the escalation events. A rule set is defined when each escalation event in the subset takes a specific outcome. Multiple risk scenarios can be associated with the same rule set. Once the individual injury probabilities of a rule set are defined, they can be used for all the associated risk scenarios. Therefore the workload for populating the risk model can be reduced by modelling on the rule sets.

There are still difficulties in modelling the uncertainty in the rule sets of an event tree containing many escalation events. First, the rule sets are intensively interwoven to

each other through the escalation events. It makes it very difficult to directly assess the correlations among rule sets. Second, the number of the rule sets increases exponentially with the number of the escalation events, which implies too many pair-wise correlations to assess for a large event tree. Third, the individual injury probabilities should have a monotonicity property: a rule set associated with worse outcomes of all the escalation events should have individual injury probabilities not less than another rule set associated with better outcomes.

According to the above difficulties, we aim to develop suitable methods for modelling the uncertainty in the rule sets of an event tree. The methods should require reasonable elicitation time from the experts. Furthermore, the methods should be able keep the monotonicity property among the rule sets.

7.1.4 Aim 3: Efficient Methods for Conducting Uncertainty Analysis of a Large Fault Trees and Event Trees

As described above, the fault trees and event trees are usually built with commercial software packages such as Isograph FT+. They are represented as the computers codes and data that are stored in the internal database of the software tools. The internal databases, however, are usually not transparent to the users in terms of the limits on accessing and manipulating the internal database. It results in difficulties in conducting uncertainty assessment in two aspects. First, for conducting simulations, we need to set the input parameters, run the model and obtain the results. It is very difficult if possible given the limits on accessing the database of the fault tree and event tree computer models. Second, the computer model composed of the codes and data stored in the internal database is completely a "black-box" to the analysts. It is impossible to do any analytical analysis with such a computer model. This is the first problem with conducting uncertainty analysis of fault trees and event trees.

For conducting the uncertainty analysis, usually we can do simulations of the model and build up the empirical distribution based on the simulations. We can also try to calculate the analytical solution of the variance of the output. For both, we need to

deal with the dependences among the input parameters, which is usually very difficult.

According to the above problems, we aim to select or develop the efficient methods to conduct uncertainty propagation through a large PRA model. The methods should be able to cooperate with the commercial software tools. The methods should also be able to work efficiently on a large Fault-tree and Event-tree model containing as many as thousands of correlated input parameters.

7.2 Summary of the Development of this Research

7.2.1 A New Procedure for Building the Covariance Matrix through the Expert Judgement Elicitation

A new procedure has been developed for building the covariance matrix through the elicitation of expert judgement. The procedure is composed of three developments including: (1) the procedure for building up the correlation matrix of a group of input parameters; (2) deriving the variance of the input parameters within a family; (3) a new way for assessing the correlation between two random variables, which is a combination of qualitative assessment and benchmark.

7.2.1.1 The Procedure for Building up the Correlation Matrix of a Group of Input Parameters

The procedure is developed based on uncertainty factors that have been employed as an efficient way to model the uncertainty in input parameters. The elicitation workload can be reduced significantly by modelling on the uncertainty factors [Van Dorp 2005]. We use uncertainty factors to structure the uncertainty of the input parameters through a linear model [Cheng, Bedford et al. 2007]. The uncertainty factors of one input parameter are assumed independent. Two uncertainty factors from different input parameters can be correlated or identical. Consequently the uncertainty factors are put into three categories: (1) exclusive factors that belong to one input parameters only; (2) common factors that belong to all the input parameters;

and (3) a factor class composed of correlated factors one and only one from each input parameter. The input parameters are correlated through the common factors and the factor classes. The correlation matrix of the input parameters is derived based on the elicitation of the correlations regarding the uncertainty factors.

An indicative value is derived to measure the proportion of the parameter's variance that is explained by the factors. This indicative value cannot exceed 1, which can be used as a constraint to monitor the assessment during the elicitation. The experts have to compare between the factors and give the consistent assessment that satisfies the constraint. As a result, a positive semi-definite correlation matrix can be guaranteed for multiple input parameters, which is a big feature of this method. Suppose there are *n* parameters having *m* common factors. The number of elicitation parameters is $n \times m$. Compared with directly filling in the correlation matrix of $n(n-1)/2$ cells, this method requires fewer values to be elicited.

7.2.1.1.1 Deriving the Variance of the Input Parameters within a Family

When a group of parameters are defined physically in a similar way, the concept of parameter family is defined. A common factor that is not affected by any differences across the family is defined as an invariant factor of the family. The set of invariant factors is called the family commonality, of which the contribution to the standard deviation of the family members is assumed to be proportional to the members' means. In the context of risk and reliability analysis, most of the input parameters are defined as the occurrence rate of the basic events [Kumamoto and Henley 1996; Modarres 2006]. The sum of the input parameters can therefore be defined as occurrence rate of the family which is meaning to assess the variance. We elicit the variance of the family from which we derive the family members' standard deviation by the invariant factors and the correlation matrix.

The method has two main features. First, the method requires less elicitation workload regarding the variance. Second, the assessment of a high level event is more reliable as agreed by the experts.

7.2.1.2 Qualitative Assessment plus Benchmark for Assessing the Correlation between Two Random Variables

As a supporting technique, we select the combination of the qualitative assessment and benchmark for eliciting the correlation between two random variables. Qualitative assessment is believed to be an easy way for experts to assess the correlation between two random variable [Clemen, Fischer et al. 2000]. The assessments however need to be mapped into numbers for conducting numerical analysis. Usually the qualitative assessments are mapped linearly into the interval between 0 and 1 [Clemen, Fischer et al. 2000]. The linear numerical mapping however implies a strong assumption and makes the qualitative assessment method less rigorous in terms of probability theory. We do the mapping in another way as employed in the Excel Add-in Crystal Ball for training the experts. For a given pair of random variables of preset correlation, the scatter plot is made on the samples and presented on the screen. The experts are then asked to draw a position on the continuous line to shown their belief of the correlation strength. By repeating this process many times, a map can be made between the positions and the underlying correlations. Based on the assumption that the experts can perceive the correlations in a way consistent with their subjective correlation perception, we think this map can be used as benchmark in the qualitative assessment of the correlations. This has proven in the case studies in this research.

The combination of the qualitative assessment and benchmark as defined above holds the threes features for a good elicitation method that are summarized in [Clemen, Fischer et al. 2000]. First, it has rigorous foundations that are defensible in terms of probability theory. Second, it is a general one that can be used in a wide variety of situations. Third, it is easy to implement and be able to be linked directly to the modelling procedure.

7.2.2 A New Method for Building the Uncertainty in the Rule Sets of an Event Tree

Along the sequence of the escalation events of an event tree, we identify the hazard types that are the materials or activities with potential to cause injuries. Depending on the outcomes of the escalation events associated with a risk scenario, each hazard type is defined at a specific level, which is called a hazard. The individual injury probabilities of the rule set are then defined in terms of the hazards.

We assume that, given the sequence, the hazards are independent in terms of causing the injuries to the people exposed to the risk scenario. When an individual person caught in a risk scenario is not lucky enough to escape all the hazards, the person becomes part of the consequence. After each hazard in the sequence, an individual exposed to the risk scenario either keeps the same level of injury or suffers a higher level injury. We assume that the transition probabilities associated with a hazard depend only on the person's injury level before the hazard. Consequently, the individual injury probabilities of a rule set can be defined on the associated hazards by a Markov Chain model [Ross 2003]. To build up the Markov Chain model, we need to elicit the transition probabilities associated with each hazard. Once the transition probabilities are elicited for one hazard, they can be used for all the rule sets containing the same hazard. Therefore, modelling on the hazards can significantly reduce the elicitation work.

For modelling the uncertainty in the rule sets, we can model the subjective uncertainty in the transition probabilities. We assume that the transition probabilities associated with different hazard types to be independent. It means that a given transition probability of one hazard does not tell any information on the transition probability of the hazards of another type. We assume that the transition probabilities associated with the same hazard type are proportional. By the assumption, an injury atom is defined for each hazard type. All the transition probabilities associated with the same hazard type are then defined on the injury atom. Consequently the individual injury probabilities of the rule sets are modelled on the injury atoms.

In most of the cases, we are more concerned with the major injuries and the fatalities than with the minor injuries. We need also to keep the elicitation work and time reasonable to the experts. For these two purposes, we develop a model of two levels of injuries as a simplified case of the model of four levels of injuries.

The above methods require reasonable elicitation time from experts. The dependences among the rule sets are modelled automatically through the injury atoms. The monotonicity property of the rule sets can also be kept between two rule sets associated with general worse and better hazards respectively.

7.2.3 Suitable Methods for Conducting Uncertainty Analysis of Fault-tree and Event-tree Models

At the first step, we develop a mimic model of the fault trees and event trees that are built with a commercial software tool. The top event of a fault tree is broken down into the basic events, also called precursors. A cut set is a collection of the basic events that together certainly cause the top event. A minimum cut set is one that is no longer a cut set when any of its basic events is removed. The basic events are called rare events when they have very small occurrence probabilities. For two minimum cut sets composed of rare events, the simultaneous occurrence probability is orders of magnitude smaller than the occurrence probability of either minimum cut sets. Based on this idea, the occurrence probability of the top event can be approximated as the sum of the occurrence probabilities of all the minimum cut sets, which is called rare event approximation [Bedford and Cooke 2001]. Once the fault trees and event trees are built with a commercial software tool such as Isograph and the input parameters are set, the minimum cut sets can be generated and associated with a risk scenario. The codes of the events composing the risk scenarios can then be output into a plain text file such as MS Excel. We can then program with MS Excel VBA on the events composing the risk scenarios to mimic the original computer model. MS Excel is a very popular software tool that offers easy access to the data. It is therefore easy to conduct simulations of the mimic model for uncertainty assessment. The mimic

model also offers a transparent structure so that we can calculate the variance of the mimic model through algebraic operations.

As assumed above, in this research we study the uncertainty in the basic events of the fault trees and the uncertainty the consequences of the event trees; we then calculate the uncertainty propagation through the fault trees and event trees. The uncertainty in the basic events and the uncertainty in the consequences are assumed to be independent. However the uncertainty in the basic events is assumed to follow a joint normal distribution. For conducting the uncertainty analysis, the basic events need to be expressed as the linear transformation of a set of independent standard normal variables. We select a linear transformation that is suitable for the context of risk analysis. Usually we implement the linear transformation through the *Cholesky* decomposition of the covariance matrix [Scheuer and Stoller 1962]. The *Cholesky* algorithm requires that the covariance matrix must be positive definite [Scheuer and Stoller 1962]. In the context of risk analysis, however, the covariance matrix can be positive semi-definite. To solve this problem, we select the linear transformation through the decomposition vectors and eigenvalues, which is called eigen-decomposition. Furthermore, in the context of risk analysis, the variance of the input parameters can be very small. The variance of the precursors related to the HET-12 in the RSSB-SRM, for example, spreads over $10^{-19} \sim 10^{-24}$ [Harrison, Griffin et al. 2008]. For such a covariance matrix, the calculation of the eigenvalues and eigenvectors is more expensive and large errors can be incurred [Wilkinson 1965; Watkins 1991]. To solve this problem, we implement the linear transformation through the eigen-decomposition of the correlation matrix instead of the covariance matrix.

Based on the linear transformation, we then develop Monte-Carlo simulations to build the empirical distribution of the output. We also develop the analytical solution for the variance through algebraic operations. The two methods are implemented independently and therefore can be used for cross-check for each other.

We then design the software to implement the above methods. The software is implemented on MS Excel workbooks and therefore can be conveniently installed and run in applications. The software is validated in three stages including: (1) validate the mimic model; (2) validate the correctness of the methods and the software; (3) validate the robustness of the methods and the software.

As a summary, the methods developed in this research can be applied directly with the computer models of the fault trees and event trees that are built with commercial software tools. The methods for assessing the uncertainty in the output is able to work efficiently even for a large model containing tens of thousands of risk scenarios.

7.2.4 Case Studies

As a validation of the developed methods, in this research we conduct two case studies of assessing the uncertainties in the output of a Fault-tree and Event-tree model.

The case studies are made on the Safety Risk Model (SRM) developed by the Rail Safety and Standards Board (RSSB). The RSSB-SRM is composed of a series of Fault-tree and Event tree models corresponding to the 125 hazardous events respectively. Currently the "best-estimated" values are set to the input parameters; consequently a point estimation of the yearly equivalent expected fatalities can be obtained from the model [Dennis 2006]. To support decision making, the assessment of the uncertainty in the output is needed. It therefore offers good case studies to test the procedure and the methods developed in this research. Our case studies are conducted on the hazardous events HET10 and HET12 as they are the two largest contributors to the safety risk in RSSB-SRM [Dennis 2006].

For the both case study, we build the covariance matrices for the precursors by the procedure developed in Chapter 3. The case studies show that the experts are satisfied to use uncertainty factors as a structure to define the correlations among a set of input parameters. The experts feel also confident to make the qualitative

assessment of the correlations regarding the uncertainty factors. In conjunction with the benchmark technique as used in *Crystal Ball* [ORACLE 2008], qualitative assessment proves an efficient way for eliciting the correlations regarding a large number of uncertainty factors. By the procedure developed in Chapter 3, the assessments of the correlations regarding the uncertainty factors are monitored during the process. The case studies show that such a monitoring scheme helps the experts to adjust their assessment. With the monitoring scheme, the outcome correlation matrices are guaranteed to be positive semi-definite. The standard deviations of the precursors are derived within the associated families. The case studies also demonstrate that the experts are confident to make the assessment on the waiting time percentile of a precursor family rather than the individual precursors.

For the both case studies, the uncertainty in the rule sets is modelled on the injury atoms by the methods developed in Chapter 4. The case studies demonstrate that the experts are confident in identifying the hazard types and structuring the rule sets in terms of the injury atoms. Modelling on the injury atoms is demonstrated as an efficient way to model the uncertainty in the rule sets including the dependence and the order relationship among the rule sets.

Based on the uncertainty models in the precursors and the rule sets built above, the uncertainty of the output is assessed by the methods developed in Chapter 6 for the both case studies respectively. The case studies demonstrate that the Excel mimic model is a reliable and efficient way to representing the Fault-tree and Event-tree models. With the case studies, the analytical solutions of the variance match the results from the simulations very well. It shows that the analytical solution is a robust way even for a large mimic model of tens of thousands of risk scenarios. The case studies also show that the corresponding software tools developed in this research are correct and reliable.

As the RSSB-SRM HET10 and HET12 represent typical Fault-tree and Event-tree models, the case studies demonstrate that the procedure and the methods developed

in Chapters 3, 4 and 5 are efficient for conducting the subjective uncertainty analysis of Fault-tree and Event-tree models.

7.3 Validation of the Development

The developments are validated in three stages. In the first stage, all the assumptions taken in this research are well discussed and agreed with the experts in RSSB. RSSB is dedicated to achieving continuous improvement in the safety performance on the Great Britain mainline railway. RSSB has built the SRM to measure the risk and the underlying causes, which has been updated to its 5th version by 2006. During the process, the experts in RSSB have gained substantial knowledge and experience in risk analysis and expert judgement elicitation. All the assumptions taken in this research have been validated by the experts in RSSB.

In the second stage, the developments based on assumptions are made based on mathematical and statistical methods and theories. The developments therefore are validated automatically.

In the third stage, we use case studies to validate the practical performance of the developments of this research. The developed procedure and methods work smoothly with the case studies. The experts are confident with making assessments following the procedure and the methods developed in this research. The methods for calculating the uncertainty propagation through the fault trees and event trees work efficiently even when tens of thousands of risk scenarios are included [Harrison, Griffin et al. 2008].

7.4 Future Research

7.4.1 The Method for Modelling the Uncertainty in the Numbers of the People Exposed to the Risk Scenarios

As discussed before, a number of people are supposed to be exposed to each risk scenario. The number needs to be elicited from the experts for each risk scenario. In

this research, the number is set as the best-estimated value. However, we cannot assess this number for sure as for the basic events and the rule sets. We therefore need to model the subjective uncertainty in the numbers of people exposed to the risk scenarios and include it in the uncertainty analysis of the fault trees and event trees.

For epistemic uncertainty analysis, we model the uncertainty in the means of the numbers of people exposed to the risk scenarios and the possible dependence. As we usually do in modelling high dimensional dependence [Kurowicka and Cooke 2006], we can assume parametrical marginal distributions for the means of the numbers and then model the dependence separately. The means, however, can have monotonicity property: the mean of the number of people exposed to one risk scenario is always larger than that to another risk scenario. It brings more difficulties in modelling the uncertainty in the means of the numbers of people exposed to the risk scenarios.

7.4.2 The Methods for Building the Uncertainty in the Outcome Probabilities of the Escalation Events

In an event tree, the escalation events can take different outcomes that affect the final consequences. Each outcome is associated with a probability that needs to be elicited from the experts. However, we can not assess the probability for sure as for others input parameters. The subjective uncertainty in the probabilities therefore needs to be modelled and included in the uncertainty analysis of the fault trees and event trees.

Typically the escalation events take two outcomes: yes or no. For such an escalation event, we can assume that probability of one outcome, say “yes”, follows a Beta distribution and the distribution can be built through expert judgement elicitation. When an escalation event has multiple outcomes, the uncertainty of the probabilities can be assumed to follow a Dirichlet distribution. We need to select or develop a method for building the distributions through expert judgement elicitation.

7.4.3 Sensitivity Analysis

After the uncertainty analysis, sensitivity analysis can be conducted to identify the most important input parameters in terms of the contributions to the uncertainty of the risk assessment model. More efforts can then be exerted to reduce the uncertainty in these important input parameters so that uncertainty of the output can be reduced most efficiently. Many methods have been developed for conducting sensitivity analysis [Saltelli, Chan et al. 2000; Saltelli 2004]. We need however to choose the suitable ones for a large risk assessment model composed of fault trees and event trees.

7.4.4 Approximate Probability Distribution of the Sum of Products of Continuous Random Variables

As discussed before, the fault trees and event trees can be approximated by the sum of products of the input parameters corresponding to the risk scenarios. For uncertainty analysis, it is desirable to derive the analytical distribution function of the output of such a model. However, the analytical probability distribution has not been developed for the product of two random variables except for some special cases.

As presented in Appendix A, we propose an approximate analytical solution of the probability distribution of the product of two random variables of a general joint distribution. Because normal random variables are used in the case studies of this research, we implement as an example the approximate probability distribution for two normal random variables. For two independent standard normal random variables, the approximate distribution is compared with the analytical distribution that has already been derived. For two correlated normal random variables, the approximate distribution is compared with the simulations.

The approximate analytical solution is further expanded for the product of multiple random variables and for the sum of the products of random variables. The random variables are dealt with sequentially and only one random variable is discretised at each step. The method is therefore suitable for large models while it is very difficult

if possible to cope with the all random variables simultaneously. We test the expanded approximate distribution model with the product of three correlated normal random variables. The approximate distribution is compared with the simulations and shows high accuracy.

The approximate analytical solutions have shown their potential to be used for uncertainty analysis of fault trees and event trees. However they are still not applicable in practice. More efforts are needed to measure and control the computation error. Once it is achieved, we can build the approximate analytical distribution function of the output of a risk assessment model composed of fault trees and event trees. The approximate probability distribution has the promising features compared with building the empirical distribution based on the simulations.

Appendix A

Approximate Probability Distribution of the Sum of Products of Continuous Random Variables

A.1 Introduction

When the rare event approximation is applied with the Fault-tree [Bedford and Cooke 2001], the Fault-tree and Event-tree models can be expressed as the sum of products of the input parameters as

$$r = f(p_1, \cdots, p_i, \cdots, p_n) = \sum_{\ell} \prod_{p_i \in R_\ell} p_i$$

where R_ℓ stands for a risk path composing of the input parameters p_i.

For probabilistic uncertainty analysis, it is desirable to obtain the analytical probability distribution of the output. Building the distribution of the product of random variables has also a wide range of applications such as radio propagation [Salo, El-Sallabi et al. 2006], Bioinformatics [Brown and Alexander 1991] and finance engineering [Nadarajah and Ali 2006] etc.

Suppose X and Y are two random variables with the joint probability distribution $f_{X,Y}(x, y)$. The distribution of $A = XY$ can be expressed through their joint distribution as

$$f_A(a) = \int_X f_{X,Y}\left(x, \frac{a}{x}\right) \cdot \frac{1}{|x|} dx$$

From this equation, the analytical solution of $f_A(a)$ can be obtained only for some specific cases such as independent $X \sim N(0,1)$ and $Y \sim N(0,1)$, and independent $X \sim U(0,1)$ and $Y \sim U(0,1)$ [Glen, Leemis et al. 2004].

Efforts have also been exerted to derive the analytical solution to $f_A(a)$ through Mellin transform [Fox 1961; Springer 1979]. The method based on Mellin transform can be extended for the product of *n* independent identical normal variables, negative exponential variables, Weibull variables, and Gamma variables respectively [Springer and Thompson 1966; Lomnicki 1967; Springer and Thompson 1970]. The method is recently extended for independent Rayleigh random variables as well [Salo, El-Sallabi et al. 2006]. The analytical solution of $f_A(a)$ can also be developed when *X* and *Y* are distributed by Lawrance Lewis's bivariate exponential distribution [Nadarajah and Ali 2006].

We can see, however, the analytical solution of $f_A(a)$ has not been derived for many other cases especially for dependent variables. In this chapter, we are going to propose a new way to build an approximate analytical solution of $f_A(a)$. The approximate analytical solution can be applied for two random variables of a generic joint distribution. The approximate solution is then extended for the product of multiple random variables and is extended further to the sum of products of random variables of a generic joint distribution. Some examples of the approximate analytical solutions are included for demonstration.

The approximate analytical solutions however are still not applicable in practice. More efforts, for example, are still needed to measure and control the computation error. We are going to show that the approximate analytical solutions potentially offer a new way to cope with a difficult task. We do not claim that these approximate analytical solutions have been soundly developed.

A.2 Mixture Distribution Model

A finite mixture distribution model is defined in [Titterington 1985; Bishop 1995; Mclachlan and Peel 2000] as

$$p(x) = \lambda_1 f_1(x) + \cdots + \lambda_K f_K(x) \tag{A.1}$$

where

$$\lambda_k > 0 \quad k = 1, \cdots, K \; ; \; \sum_{k=1,K} \lambda_k = 1$$

$$f_k(x) \geq 0 \, , \; \int_X f_k(x) dx = 1$$

Following [Titterington 1985; Bishop 1995; Mclachlan and Peel 2000], the parameters λ_k are called mixing weight and $f_k(x)$ the component densities of the mixture.

When the all component densities belong to the same parametric family, the mixture distribution model defined in Equation A.1 becomes

$$p(x) = \sum_{k=1,K} \lambda_k \cdot f(x|\theta_k) \tag{A.2}$$

where θ_k denotes the parameter occurring in the k^{th} component associated with λ_k.

The mixture weights $\lambda = (\lambda_1, \cdots, \lambda_K)$ can also be thought of as a discrete prior distribution on the parameters θ_k [Titterington 1985]. When the parameter θ is defined on a continuous domain, the mixture weight is defined by the weight distribution $\lambda(\theta)$ that satisfies

$$\lambda(\theta) > 0 \; ; \; \int_{\theta} \lambda(\theta) d\theta = 1$$

Correspondingly a continuous mixture distribution model can be defined as in [Gelman, Carlin et al. 2003] as

$$p(x) = \int_{\theta} f(x|\theta) \lambda(\theta) d\theta \tag{A.3}$$

From the above mixture distribution models, an analytical solution of the distribution of the product of random variables is to be derived.

A.3 Distribution of Product of Two Random Variables

A.3.1 Mixture Distribution Model for Product of Two Random Variables

Suppose X and Y are two continuous random variables with the joint distribution $f_{X,Y}(x, y)$. The joint distribution can be expressed as [Ross 2003]:

$$f_{X,Y}(x, y) = f_X(x) \cdot f_{Y|X=x}(y) \tag{A.4}$$

where $f_X(x)$ is the marginal distribution of X and $f_{Y|X=x}(y)$ is the distribution of Y conditional on $X = x$.

Suppose

$$A = X \cdot Y \tag{A.5}$$

We need to derive the distribution of A denoted as $f_A(a)$. From Equation A.5, the distribution of A conditional on $X = x \neq 0$ can be derived by standard transformation of variables as

$$f_{A|X=x}(a) = \frac{1}{|x|} f_{Y|X=x}\left(\frac{a}{x}\right) \tag{A.6}$$

We can then treat $f_{A|X=x}(a)$ as the component and $f_X(x)$ the weight density of a mixture distribution model as defined in Equation A.3. Subsequently, $f_A(a)$ can be defined as a mixture distribution model as

$$\begin{aligned} f_A(a) &= \int_X f_{A|X=x}(a) \cdot f_X(x) dx \\ &= \int_X \frac{1}{|x|} f_{Y|X=x}\left(\frac{a}{x}\right) \cdot f_X(x) dx \end{aligned} \tag{A.7}$$

Equation A.7 defines the distribution function $f_A(a)$ as integral over *X*. For some types of $f_{Y|X=x}(y)$, a closed form of $f_A(a)$ can be derived from the integration. For general conditional distribution $f_{Y|X=x}(y)$, the integral can be approximated with the summation based on a discretisation of *X*. We have shown that $f_{A|X=x}(a)$ is a parametric function of *a*, which can be derived from the parametric function $f_{Y|X=x}(y)$ given $X = x$ by Equation A.6. An approximate analytical function to $f_A(a)$ will be derived as the sum of the parametric functions of the form $f_{A|X=x}(a)$ accordingly.

A.3.2 Approximate Distribution Function

Based on Equation A.7, the support of *X* can be partitioned into *K* intervals by setting $K+1$ borders $x_{B,k}$, $k = 0, \cdots, K$, where $x_{B,0} < \cdots < x_{B,k} < \cdots < x_{B,K}$. The two terminal borders $x_{B,0}$ and $x_{B,K}$ are set as the bounds of *X*. When *X* is defined on $(-\infty, +\infty)$, $x_{B,0}$ and $x_{B,K}$ can be defined by

$$F_X(x_{B,0}) = TPB, \; F_X(x_{B,K}) = 1 - TPB$$

where $F_X(x)$ represents the cumulative probability of *X*; *TPB* represents the truncated probability bound and should be a small positive real number.

Denote the interval $\lfloor x_{B,k-1}, x_{B,k} \rfloor$ as $X_{I,k}$. The probability of $x \in X_{I,k}$ is defined as

$$m_k = \Pr(x \in X_{I,k}) = F_X(x_{B,k}) - F_X(x_{B,k-1}) \tag{A.8}$$

For each interval $X_{I,k}$, a point $x_k \in X_{I,k}$ can be chosen as the representative. For instance, the representative can be simply set as the middle point as

$$x_k = (x_{B,k-1} + x_{B,k})/2$$

The continuous random variable *X* is now discretised as

$$\Pr(X = x_k) = m_k, \; k = 1, \cdots, K \tag{A.9}$$

We have shown that $f_{A|X=x}(a)$ is a parametric function of *a*, which can be derived from the parametric function $f_{Y|X=x}(y)$ given $X = x$ by Equation A.6. Corresponding to the discretisation of *X*, the component $f_{A|X=x}(a)$ can be approximated with $f_{A|X=x_k}(a)$ for $x \in X_k$. The integral in Equation A.7 can then be approximated with the following summation

$$g_A(a) = \sum_{k=1,K} m_k f_{A|X=x_k}(a) = \sum_{k=1,K} m_k \cdot \frac{1}{|x_k|} f_{Y|X=x_k}\left(\frac{a}{x_k}\right) \tag{A.10}$$

When the parametric function $f_{Y|X=x}(y)$ is known, the distribution function $g_A(a)$ is defined by Equation A.10 as an approximation to $f_A(a)$ that itself cannot be derived explicitly for *X* and *Y* of a general joint distribution. The definition of $g_A(a)$ is demonstrated in Fig. A.1. We can see that $g_A(a)$ is an approximate analytical

solution that is a function of a. It therefore is different from the traditional numerical integral that calculates a discrete probability density of A =a. Therefore the traditional integration methods as summarized in [Burden and Faires 1997] are not suitable to generate the approximate distribution $g_A(a)$ defined in Equation A.10 from the original integral defined in Equation A.11. Based on the mixture probability model defined in Equations A.3 and A.2, we actually propose a way to discretised X by the marginal distribution $f_X(x)$, the conditional distribution $f_{Y|X=1}(y)$ and the inverse curve $1/x$. We are going to discuss the method for discretising X in more details.

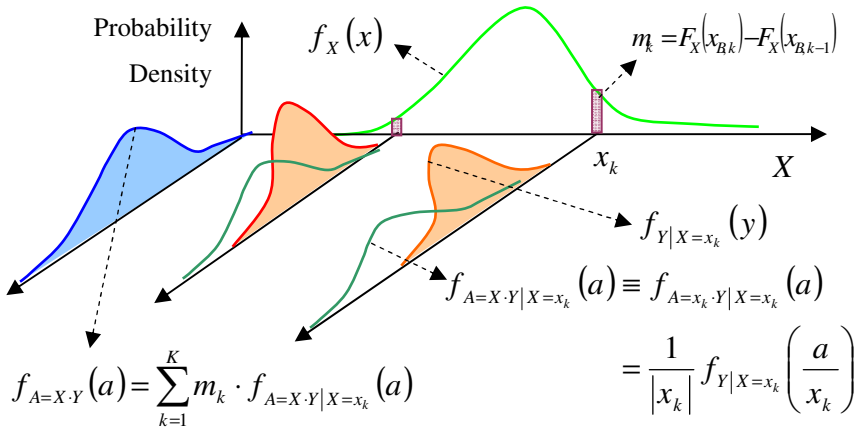


Fig. A.1. The illustration for building the approximate parametric function of the distribution of the product of two continuous random variables

By comparing Equation A.7 and A.10, we can see that the computation errors are incurred when partitioning X into the intervals $X_{I,k}$. Define $\Delta x_k = x_{B,k} - x_{B,k-1}$. Firstly the component weight density in Equation A.7 is defined by $f_X(x)$ while it is approximated with $m_k/\Delta x_k$ for $x \in X_k$ in Equation A.10. Secondly the scaling coefficient in Equation A.7 is defined by $1/x$; while it is approximated with $1/x_k$ for $x \in X_k$ in Equation A.10. Thirdly the conditional distribution in Equation A.7 is

defined by $f_{Y|X=x}(y)$; while it is approximated with $f_{Y|X=x_k}(y)$ for $x \in X_k$ in Equation A.10. By properly setting $x_{B,0}$, $x_{B,K}$ and x_k for $k = 1, K$ the variability of $f_X(x)$, $1/x$ and $f_{Y|X=x}(y)$ can be controlled for $x \in X_{I,k}$. Subsequently the overall error can be controlled within an acceptable level. Some general principles include:

- Keep smaller Δx_k where there is a larger absolute value for $\frac{d}{dx} f_X(x)$, $\frac{d}{dx}(1/x)$ or $\frac{\partial}{\partial y}\frac{\partial}{\partial x} f_{X,Y}(x, y)$;
- When the jump discontinuity points of $f_X(x)$ and $f_{X,Y}(x, y)$ exist, they have to be set as a border $x_{B,k}$;
- For unbounded *X*, the borders $x_{B,0}$ and $x_{B,K}$ must be set to ensure the probability $\Pr(X \in \lfloor x_{B,0}, x_{B,K} \rfloor)$ is large enough and the impact of the truncated probability has to be accounted properly.
- A singular point of Equation A.7 exists at $X = 0$. In case of $0 \in \lfloor x_{B,0}, x_{B,K} \rfloor$, three successive borders must be set at $-\delta$, 0 and δ where $\delta > 0$ is a small real number specified for the expected accuracy.

We are going to test the above methods by building the distribution of two normal random variables because they are used popularly for many applications including RSSB-SRM on which the case studies are to be carried in this research. We will test the impact of the different settings of *TPB*, $x_{B,k}$, x_k on the approximation accuracy of the mixture model defined in Equation A.10.

A.3.3 Distribution of the Product of two Normal Random Variables

A.3.3.1 Approximate Probability Distribution Function

The analytical solution of the distribution of the product of two independent standard normal variables can be built based on Bessel function [Weisstein ; Springer 1979; Glen, Leemis et al. 2004]. For two general normal random variables, however, no

analytical solution has been found in the literature. We develop the approximate distribution function of the product of two general normal variables by Equation A.10. This approximate function is then implemented with two independent standard normal variables, and two correlated normal variables respectively.

Suppose that X and Y are two normal random variables denoted as $X \sim N(\mu_X, \sigma_X)$ and $Y \sim N(\mu_Y, \sigma_Y)$, where μ_X and μ_Y are the mean of X and Y respectively; σ_X and σ_X are the standard deviation of X and Y respectively. Denote the correlation between X and Y as $r(X,Y)$. The joint distribution of X and Y is completely defined by μ_X, μ_Y, σ_X, σ_X and $r(X,Y)$.

The conditional distribution of Y given $X = x$ is still a normal distribution that is defined as given in [Tong 1990] as

$$f_{Y|X=x}(y) = N\left(\mu_{Y|X=x}, \sigma_{Y|X=x}\right) \tag{A.11}$$

where

$$\mu_{Y|X=x} = \mu_Y + r(X,Y)\frac{\sigma_Y}{\sigma_X}(x - \mu_X) \tag{A.12}$$

$$\sigma_{Y|X=x} = \sigma_Y\sqrt{1 - r(X,Y)} \tag{A.13}$$

Based on Equation A.5, the product A given $X = x$ is normally distributed as well [Ross 2003]. The conditional distribution of A in Equation A.6 can be defined as

$$f_{A|X=x}(a) = N\left(\mu_{A|X=x}, \sigma_{A|X=x}\right) \tag{A.14}$$

where

$$\mu_{A|X=x} = x \cdot \mu_{Y|X=x} = x \cdot \left(\mu_Y + r(X,Y)\frac{\sigma_Y}{\sigma_X}(x - \mu_X) \right) \tag{A.15}$$

$$\sigma_{A|X=x} = |x| \cdot \sigma_{Y|X=x} = |x| \cdot \sigma_Y \sqrt{1 - r(X,Y)} \tag{A.16}$$

Based on Equation A.10 the approximate distribution of *A* can be derived as

$$g_A(a) = \sum_{k=1,K} m_k \cdot N\left(\mu_{A|X=x_k}, \sigma_{A|X=x_k}\right) \tag{A.17}$$

Equation A.17 defines the approximate analytical distribution of the product of two normal random variables as a mixture model of a set of normal distributions. Three examples are made by using the software package MATLAB to demonstrate the performance of this mixture model. The simulation samples of *X* and *Y* are generated directly by using MATLAB function *mvnrnd()*

A.3.3.2 Example 1: Uncorrelated Standard Normal *X* and *Y*

Suppose that *X* and *Y* are standard normal random variable denoted as $X \sim N(0,1)$ and $Y \sim N(0,1)$. Suppose also $r(X,Y) = 0$. The analytical solution of $f_A(a)$ has been derived based on Bessel function in [Weisstein; Springer 1979; Glen, Leemis et al. 2004] as

$$f_A(a) = \frac{K_0(|a|)}{\pi}, \; a \neq 0 \tag{A.18}$$

where $K_0(\,)$ is a modified Bessel function of the second kind.

We are going to derive the approximated distribution $g_A(a)$ as defined in Equation A.17. We then compare the approximate distribution with the analytical solution.

Because *X* and *Y* are independent, we have

$$\mu_{Y|X=x} = \mu_Y = 0$$

$$\sigma_{Y|X=x} = \sigma_Y = 1$$

As defined in Equations A.15 and A.16, the component mean and standard deviation are calculated as

$$\mu_{A|X=x} = x \cdot \mu_{Y|X=x} = 0$$

$$\sigma_{A|X=x} = |x| \cdot \sigma_{Y|X=x} = |x|$$

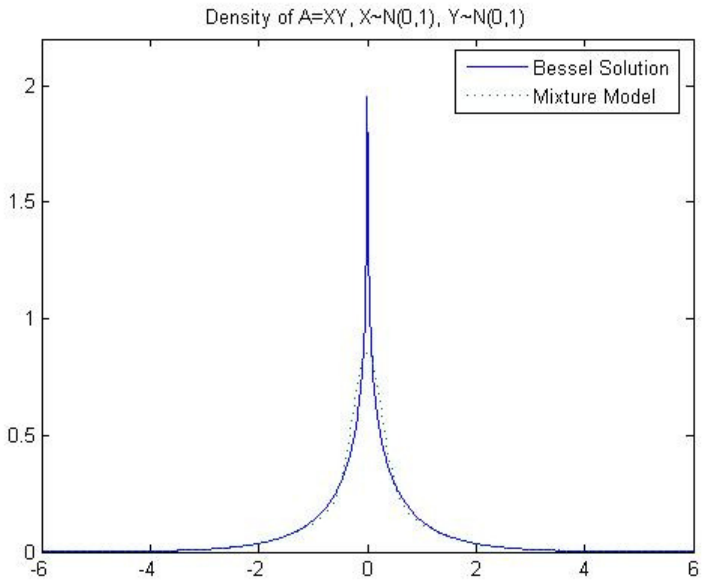
We start with setting the *TPB* as 1e-8 and discretise *X* by keeping the same value of $|f_X(x_{B,k}) - f_X(x_{B,k-1})|$. The representative point is set as $x_k = (x_{B,k} - x_{B,k-1})/2$. Based on Equation A.17, two mixture models are built with 16 and 200 components for comparison. The weight, the mean and the standard deviation of the 16 components are defined in Table A.1. For this example, we can see that the components are symmetrically defined corresponding to the symmetric samples of *X*. Therefore the approximate distribution can actually be defined by 8 components and 100 components respectively.

Table A.1 Normal-components of the approximate distribution of the product of two independent standard normal random variables

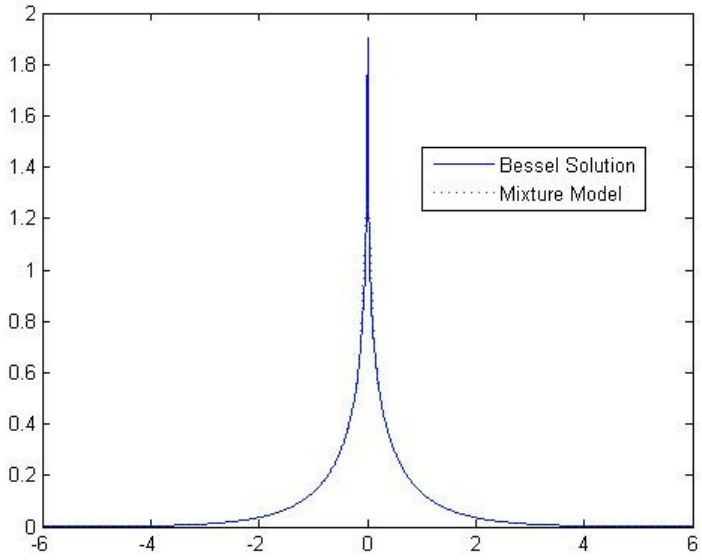
No.	Mixing Weight	Component Mean	Component Std
1.	5.00E-11	0	12.934
2.	0.024261	0	3.2891
3.	0.032463	0	1.7462
4.	0.039774	0	1.4284
5.	0.048544	0	1.1712
6.	0.060971	0	0.93255

7.	0.083349	0	0.68178
8.	0.21064	0	0.26696
9.	0.21064	0	0.26696
10.	0.083349	0	0.68178
11.	0.060971	0	0.93255
12.	0.048544	0	1.1712
13.	0.039774	0	1.4284
14.	0.032463	0	1.7462
15.	0.024261	0	3.2891
16.	5.00E-11	0	12.934

For comparison, the 8-component approximate distribution and the 100-component approximate distribution are plotted together with the analytical distribution in Fig. A.2 (a) and (b) respectively. To demonstrate the difference further, we plot the ratio of the analytical distribution to the approximate distribution, i.e. $f_A(a) \big/ g_A(a)$, in Fig. A.3 (a) and (b) for the two approximate distributions respectively. The plots demonstrate that the approximate distributions coincides with the analytical distribution very well except in a small area around $A = 0$. The reason is that the approximate distribution has a finite density value for $A = 0$ while the analytical distribution has an infinite definition for $A = 0$. We can however gain higher accuracy by including more components in the approximate distribution.

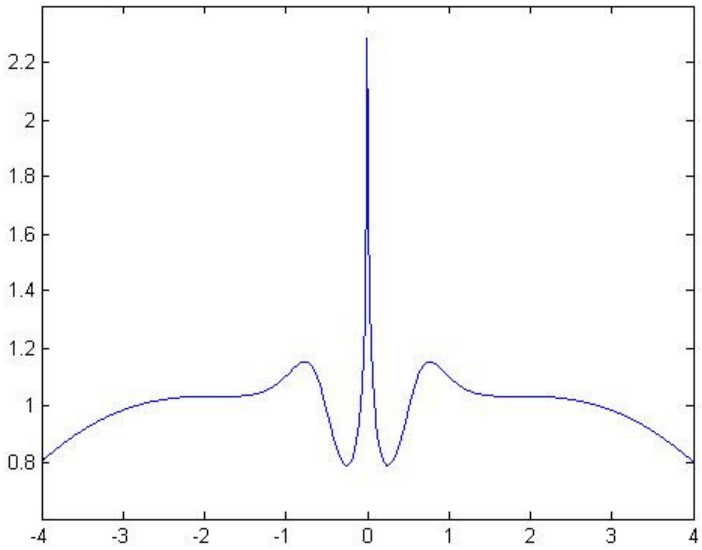


(a) for the approximate distribution of 8 components

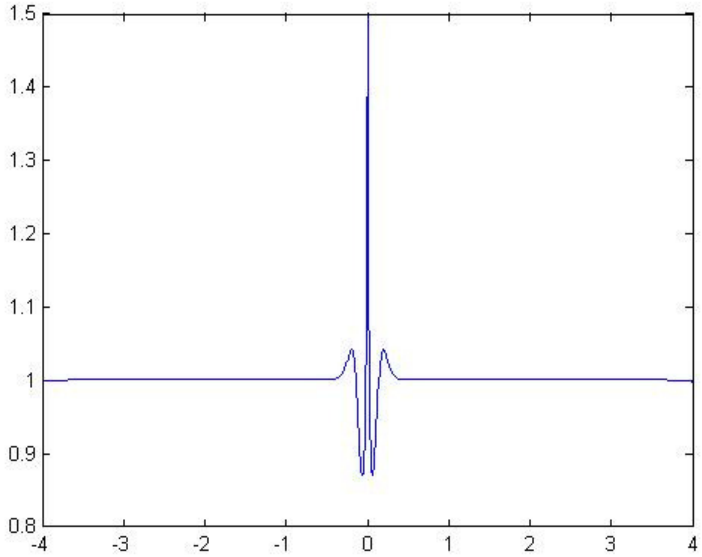


(b) for the approximate distribution of 100 components

Fig. A.2 The analytical distribution and the approximated distribution of the product of two independent standard normal random variables



(a) for the approximate distribution of 8 components



(b) for the approximate distribution of 100 components

Fig. A.3 The ratio between the analytical distribution and the approximate distribution of the product of two independent standard normal random variables

A.3.3.3 Example 2: Uncorrelated non-standard Normal *X* and *Y*

We design this example to demonstrate the performance of the approximate distribution of the product of two independent normal variables that have non-zero means. We consider X~$N(12, 0.4)$, Y~$N(3, 0.2)$ and $r(X,Y)=0$. For such a case, the analytical distribution $f_{A=XY}(a)$ has not been found in the literature. The analytical mean and the standard deviation however can be calculated by

$$E(A)=E(XY)=E(X)E(Y)=36$$

$$\mathrm{var}(A)=\mathrm{var}(XY)=\mathrm{var}(X)\cdot E(Y)^2+\mathrm{var}(Y)\cdot E(X)^2+\mathrm{var}(X)\cdot\mathrm{var}(Y)$$

$$\sigma(A)=\sqrt{\mathrm{var}(A)}=2.68447387769$$

We define the relative errors as

$$\mathrm{err}_{\mathrm{E}}=\left(E(A)\big|_{g_A(a)}-E(A)\right)\Big/E(A)$$

$$\mathrm{err}_{\sigma}=\left(\sigma(A)\big|_{g_A(a)}-\sigma(A)\right)\Big/\sigma(A)$$

where $E(A)\big|_{g_A(a)}$ and $\sigma(A)\big|_{g_A(a)}$ stand for the mean and standard deviation that are calculated from the approximate distribution $g_A(a)$.

The relative errors are used to measure the performance of the approximate distribution model with different implementation settings. As summarised in Table A.2, three implementation cases are defined with the combinations of the ways to delimit the borders $x_{B,k}$ and the ways to treat the truncated probability. For all the three cases, the representative x_k is set by equally dividing m_k, i.e. by satisfying

$$F_X(x_{B,k})-F_X(x_k)=F_X(x_k)-F_X(x_{B,k-1})=m_k/2$$

For all the three cases, the random variable *X* is discretised into 300, 500, 800, 1000, 1600 and 2000 points respectively; and *TPB* is set as 1e-5, 1e-6, 1e-8, and 1e-10 respectively.

Table A.2. The cases of Example 1 for building the distribution of *A=XY*

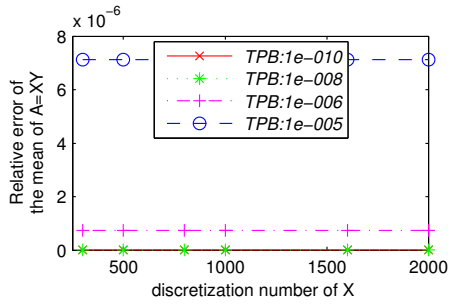
	Treatment of the truncated probabilities	**Delimiting borders** $x_{B,k}$, for $k = 1 \cdots K-1$
Case 1 Fig. A.2 (a), (b)	Attributed to a representative $x_{B,0}/2$ for $[-\infty, x_{B,0}]$ and $2x_{B,K}$ for $[x_{B,K}, +\infty]$ respectively.	Set $x_{B,k}$ by equally dividing the density curve, i.e. keep constant $\left\|f_X(x_{B,k}) - f_X(x_{B,k-1})\right\|$
Case 2 Fig. A.2 (c), (d)	Attributed to a representative $x_{B,0}/2$ for $[-\infty, x_{B,0}]$ and $2x_{B,K}$ for $[x_{B,K}, +\infty]$ respectively.	Set $x_{B,k}$ by equally dividing the inverse curve, i.e. keep constant $\frac{1}{x_{B,k-1}} - \frac{1}{x_{B,k}}$
Case 3 Fig. A.2 (e), (f)	Attributed to x_1 for $[x_{B,0}, x_{B,1}]$ and x_K for $[x_{B,K-1}, x_{B,K}]$ respectively	Set $x_{B,k}$ by equally dividing the inverse curve, i.e. keep constant $\frac{1}{x_{B,k-1}} - \frac{1}{x_{B,k}}$

In Fig. A.4 (a)-(f), we plot the relative errors against *TPB* and the discretisation number for the three cases respectively. For all the three cases, it shows that *TPB* and the discretisation number can change the accuracy subject to the pattern defined by the other two factors. The minimum relative errors given under the diagrams demonstrate that the approximate distribution can achieve very high accuracy with affordable computation load.

We now compare the performance of the three cases based on the relative errors that are illustrated in Fig. A.4. By comparing Fig. A.4 (a) and Fig. A.4 (c), we can find that Case 1 shows higher accuracy on the mean of *A*. By comparing Fig. A.4 (b) and Fig. A.4 (d), however, we can find that Case 1 shows poorer accuracy on the standard deviation of *A* than Case 2. As summarized in Table A.2, Case 1 and Case 2 have the same way to attribute the truncated probabilities. The borders $x_{B,k}$ are set by uniformly dividing the density curve in Case 1; while the borders $x_{B,k}$ are set by uniformly dividing the inverse curve in Case 2. Therefore we cannot see either way for setting the borders $x_{B,k}$ in Case 1 and Case 2 is better.

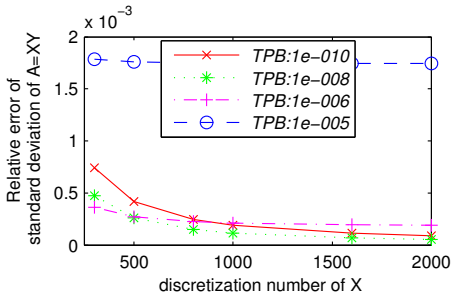
By comparing Fig. A.4 (c) with Fig. A.4 (e) and comparing Fig. A.4 (d) with Fig. A.4 (f), we can see that Case 2 has higher accuracy on both the mean and the standard deviation of *A* than Case 3 does. Case 2 and Case 3 have the same way to set the borders $x_{B,k}$. In Case 2 the truncated probabilities are attributed to two extra samples $x_{B,0}/2$ and $2x_{B,K}$ on the both sides respectively. In Case 3, however, the truncated probabilities are attributed to the two terminal samples x_1 and x_K respectively. It shows that the treatment of the truncated probabilities can make a significant difference and Case 2 represent a set of implementation settings of better performance.

Fig. A.5 illustrates the approximate distribution $g_A(a)$ that is built by Case 2 with 2000 discretisation points of *X* and *TPB* set as 1e-10. The Q-Q plot of $g_A(a)$ against 60000 simulations of *A* is also drawn on Fig. A.5. We can see that the Q-Q plot coincides with the Q=Q line very well. It shows that $g_A(a)$ is a good approximation to the real distribution $f_A(a)$. The relative errors of the mean and standard deviation are about 8.0e-9 and 7.4e-7 respectively.



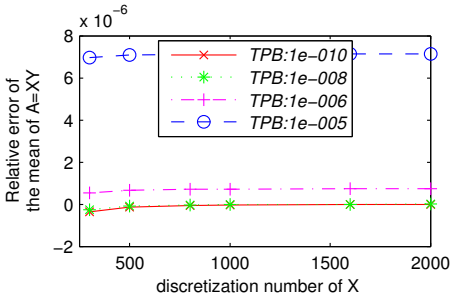
(a) - Case 1, mean

(Minimum relative error 8.2e-11)



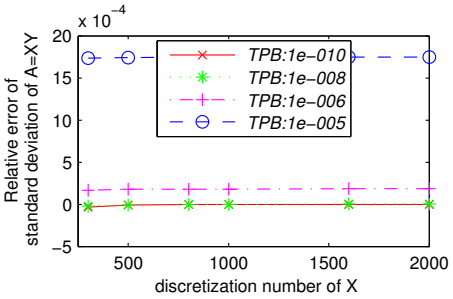
(b) Case 1, standard deviation

(Minimum relative error 5.1e-5)



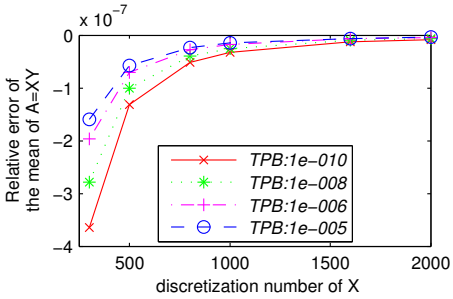
(c) - Case 2, mean

(Minimum relative error 1.5e-9)



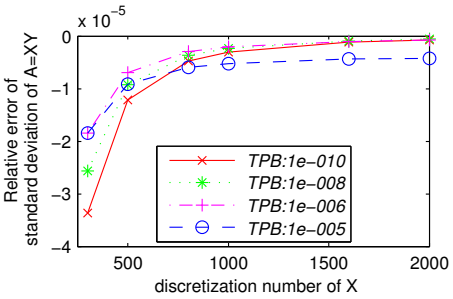
(d) - Case 2, standard deviation

(Minimum relative error 3.6e-7)



(e) - Case 3, mean

(Minimum relative error 3.9e-9).



(f) - Case 3, standard deviation

(Minimum relative error 5.8e-7)

Fig. A.4. The illustration of the relative errors of Example 1. The cases are defined in Table A.2.

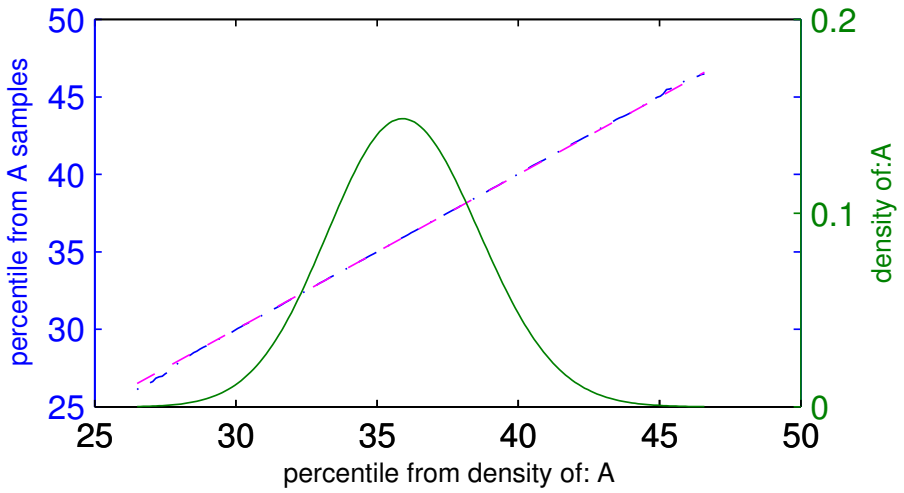


Fig. A.5. This diagram illustrates the approximate probability distribution of $A=XY$ where X~N(12, 0.4), Y~N(3, 0.2) and $r(X,Y)=0$. The green solid curve represents the approximate probability distribution $g_A(a)$ that is built from Case 2 in Table A.2. The blue dash-dot curve represents the Q-Q plot against the 60000 simulations of A. The magenta dash line stands for Q=Q line.

A.3.3.4 Example 3: Strongly Correlated *X* and *Y*

In this example, we test the performance of the approximate probability distribution $g_A(a)$ for correlated normal variables X and Y. As in Example 2, we set X~N(12, 0.4) and Y~N(3, 0.2). We set the correlation as $r(X,Y)=0.99999$. The borders $x_{B,k}$, the representative x_k and the truncated probabilities are set in the same as in Case 2 of Example 2. The *TPB* is set as 1e-10. We build the approximate probability distribution $g_A(a)$ with 600 and 3000 discretisation points of X respectively. The probability distributions are plotted in Figs A.6 (a) and (b) respectively. The Q-Q plots against 60000 simulations of A are added in Fig. A.6 as well.

As illustrated in Fig. A.6 (a), the approximate probability distribution corresponding to 600 discretisation points of X shows lots of spikes; and correspondingly the Q-Q

plot demonstrates lots of very small waves apart from the Q=Q. Comparatively, as illustrated in Fig. A.6 (b), the approximate probability distribution corresponding to 3000 discretisation points of *X* is very smooth and the Q-Q plot shows little diversion from the Q=Q line. It implies huge error of the approximate probability distribution corresponding to 600 discretisation points of *X*. We will show that the reason is due to the strong correlation between *X* and *Y*, which causes large change in $f_{Y|X=x}(y)$ for $x \in \left[x_{B,k-1}, x_{B,k}\right]$.

The bivariate joint distribution of *X* and *Y* is defined as

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}} \exp\left(\frac{-1}{2\left(1-\rho_{xy}^2\right)}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}-\frac{2\rho_{xy}xy}{\sigma_x\sigma_y}\right)\right) \quad \text{(A.19)}$$

where $\rho_{x,y}$ is the correlation coefficient between *X* and *Y*.

The partial derivative of the joint distribution is derived as

$$\begin{aligned}\frac{\partial}{\partial y}\frac{\partial}{\partial x} f_{X,Y}(x,y) &= \frac{1}{\sigma_x\sigma_y}\cdot\frac{\rho_{x,y}}{1-\rho_{x,y}^2}\cdot f_{X,Y}(x,y) \\ &= \frac{1}{0.2\times 0.4}\cdot\frac{0.99999}{1-0.99999^2}\cdot f_{X,Y}(x,y) \\ &= 6.25\times 10^5 \cdot f_{X,Y}(x,y)\end{aligned} \quad \text{(A.20)}$$

Equations A.20 shows that the derivative is more than 6.25×10^5 times the density when $r(X,Y) = 0.99999$. It means huge variability of $f_{Y|X=x}(y)$ and requires high discretisation number of *X* as discussed previously. This fact can be demonstrated by Equations A.12 and A.13 as well, from which we have

$$\begin{aligned}\mu_{A|X=x} &= x\cdot\left(\mu_Y + r(X,Y)\frac{\sigma_Y}{\sigma_X}(x-\mu_X)\right) \\ &= x\cdot(\mu_Y + 0.5(x-\mu_X))\end{aligned}$$

$$\sigma_{A|X=x} = |x| \cdot \sigma_Y \sqrt{1 - r(X,Y)} = 6.32 \times 10^{-4} \cdot |x|$$

When *X* changes from x to $x + \Delta x$, $\Delta x > 0$, we have the offset

$$\mu_{A|X=x+\Delta x} - \mu_{A|X=x} = 0.5(\Delta x)(\Delta x) + (x - 3)\Delta x$$

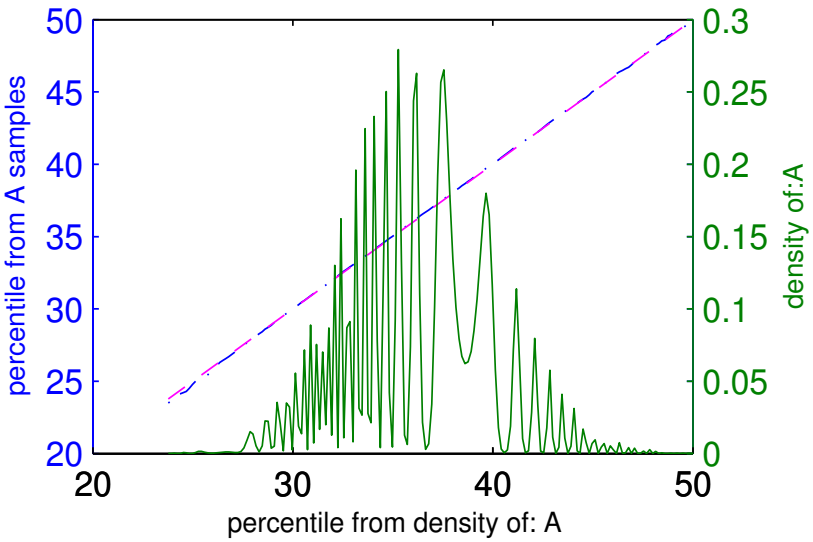
When we set *TPB* as 1e-10, we have $x_{B,0} = 9.4$, $x_{B,K} = 14.6$. Because $x_{B,0} < x < x_{B,K}$, we have

$$\sigma_{A|X=x} < 9.3 \times 10^{-3}$$

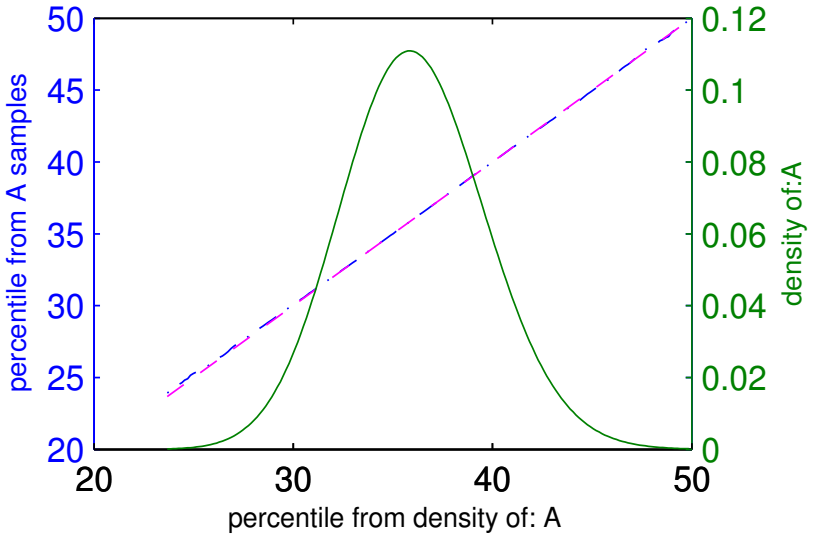
$$\mu_{A|X=x+\Delta x} - \mu_{A|X=x} > 6.4 \cdot \Delta x$$

When we use 600 discretisation points of *X*, the average width Δx is $(x_{B,K} - x_{B,0})/600 = 8.7 \times 10^{-3}$. Therefore we have $\mu_{A|X=x+\Delta x} - \mu_{A|X=x} > 5.5 \times 10^{-2}$, which is 5.9 time the $\sigma_{A|X=x}$. As defined in Equation A.14, $f_{A|X=x}(a)$ is the probability distribution of a normal variable of the mean $\mu_{A|X=x}$ and the standard deviation $\sigma_{A|X=x}$. Therefore a huge error is encountered when we use $f_{A|X=x_k}(a)$ as a representative of $f_{A|X=x}(a)$, for $x \in [x_{B,k-1}, x_{B,k}]$. It explains why the approximate probability distribution built with 600 discretisation points of *X* shows many spikes in Fig A.6 (a). When we increase the discretisation points of *X* to 3000, we have $\mu_{A|X=x+\Delta x} - \mu_{A|X=x} > 1.1 \times 10^{-2}$, which is 1.17 time the $\sigma_{A|X=x}$. The error in representing $f_{A|X=x}(a)$, for $x \in [x_{B,k-1}, x_{B,k}]$ by $f_{A|X=x_k}(a)$ is therefore much less than that with 600 discretisation points of *X*. A smooth approximate probability distribution is then obtained as shown in Fig A.6 (b).

As a summary, this example demonstrates that the approximate probability distribution as defined in Equation A.17 is able to deal with the strongly correlated normal random variables. It shows also that we need to much more discretisation points of X for strongly correlated X and Y to achieve acceptable accuracy. More generally it exemplifies the principle for discretising X regarding a large derivative of the joint distribution, i.e. $\frac{\partial}{\partial y}\frac{\partial}{\partial x} f_{X,Y}(x, y)$.



(a) for 600 discretisation points of X



(b) for 3000 discretisation points of X

Fig. A.6. These diagrams illustrate the approximate probability distributions of $A=XY$, where X~$N(12,\ 0.4)$, Y~$N(3,\ 0.2)$ and $r(X,Y)=0.99999$. The green solid curve represents the approximate probability distribution. The blue dash-dot curve represents the Q-Q plot against 6000 simulation of A. The magenta dash line stands for Q=Q line.

A.4 Approximate Probability Distribution of the Product of Multiple Random Variables

A.4.1 A Sequential Process

A product of multiple random variables is defined as

$$\prod_{i=1,n} X_i = X_1 \cdots X_n \tag{A.21}$$

where $n > 2$; $X_1, \cdots, X_n$ are random variables.

We suppose that the joint distribution of $X_1, \cdots, X_n$ is defined as

$$f_{X_1,\cdots X_n}(x_1, \cdots, x_n) = f_{X_1}(x_1) \cdot f_{X_2|X_1}(x_2) \cdots f_{X_n|X_{n-1},\cdots X_1}(x_n) \tag{A.22}$$

For reference convenience, the right sides of Equations A.21 and A.22 are called random variable chain (RVC) and the associated conditional distribution chain (CDC) respectively. As illustrated in Fig. A.7, the RVC and the associated CDC are put side by side to define the product. We start from the first row. The idea is to replace X_1 and X_2 with $A_2 = X_1 X_2$ that is called surrogate. For this purpose, we need to build the approximate marginal distribution $g_{A_2}(a_2)$ and, correspondingly, to update the CDC with $g_{A_2}(a_2) \cdot g_{X_3|A_2}(x_3) \cdot \cdots \cdot g_{X_n|X_{n-1}\cdots X_3,A_2}(x_n)$. After finishing these two tasks,

the original product $\prod_{i=1,n} X_i$ can be approximated by $A_2 \cdot \prod_{i=3,n} X_i$ as defined on the second row. We then continue going down by defining $A_i = A_{i-1} X_i$ for $i = 3, \cdots, n$ and repeating the above process. The last surrogate A_n approximates the original product $\prod_{i=1,n} X_i$. To implement the algorithm, the approximate marginal distribution $g_{A_i}(a_i)$ needs to be built firstly, which can be implemented with the mixture model for two random variables as defined in Equation A.10. Secondly the associated CDC needs to be updated with A_i replacing A_{i-1} and X_i at each stage, for which a new algorithm is to be developed based on Bayes' theorem.

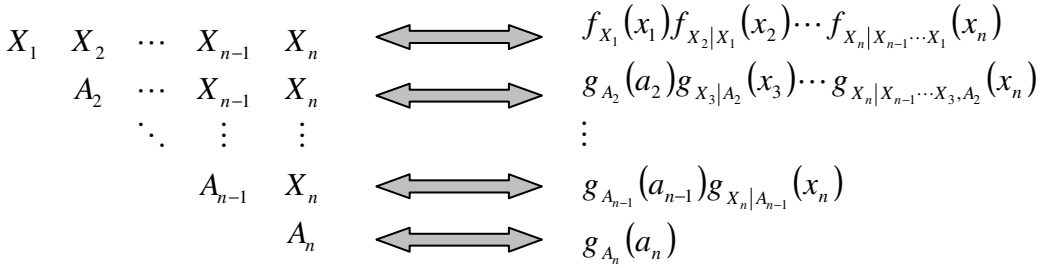


Fig. A.7. The illustration of the process for building the approximate probability distribution of the product of multiple random variables $X_1, \cdots, X_n$.

A.4.2 Method for Updating the Conditional Distribution Chain

Following the definitions given by Equations A.21 and A.22, the first surrogate *A* is defined as

$$A = X_1 X_2 \tag{A.23}$$

By substituting Equation A.23 with Equation A.21 we get a new RVC $A_1 X_3 \cdots X_n$. We will build the CDC for $A_1 X_3 \cdots X_n$ based on the CDC defined in Equation A.22.

The random variable X_1 is discretised as defined in Equation A.9 as

$$\Pr(X_1 = x_{1,k}) = F_{X_1}(x_{B1,k}) - F_{X_1}(x_{B1,k-1}) = m_{1,k}$$

where $x_{1,k}$ and $m_{1,k}$ stand for a discrete point and the associated mess probability of X_1 respectively.

The approximate distribution $g_A(a)$ can then be built as a mixture model by Equation A.10. By Bayesian theorem [Bishop 1995], the probability of $X_1 = x_{1,k}$ given $A = a$ can be derived as

$$\Pr(X_1 = x_{1,k} | A = a) = \frac{m_{1,k} \cdot f_{A|X_1 = x_{1,k}}(a)}{g_A(a)} \tag{A.24}$$

As discussed previously, three successive borders are set at $-\delta$, 0 and δ, where $\delta > 0$, to avoid the singular point $x_{1,k} = 0$ of Equation A.10. As a result, it is guaranteed that $x_{1,k} \neq 0$, $\forall k = 1, \cdots, K$. Subsequently based on Equation A.24 we have

$$\Pr\left(X_2 = \frac{a}{x_{1,k}} \middle| A = a, X_1 = x_{1,k} \right) = 1 \tag{A.25}$$

Based on Equation A.24-A.25, we have

$$\Pr\left(X_1 = x_{1,k}, X_2 = \frac{a}{x_{1,k}} \middle| A = a \right) = \frac{m_{1,k} \cdot f_{A|X_1 = x_{1,k}}(a)}{g_A(a)} \tag{A.26}$$

The conditional distribution $f_{X_3|A=a}(x_3)$ can be approximated as a mixture model as defined in Equation A.2 as

$$g_{X_3|A=a}(x_3)=\sum_{k=1,K}\frac{m_{1,k}f_{A|X_1=x_{1,k}}(a)}{g_A(a)}\cdot f_{X_3\left|X_2=\frac{a}{x_{1,k}},X_1=x_{1,k},\right.}(x_3) \tag{A.27}$$

By the same way, we have the approximate conditional distribution defined as

$$g_{X_n|X_{n-1}=x_{n-1},\cdots,X_3=x_3,A=a}(x_n)=\sum_{k=1,K}\frac{m_{1,k}f_{A|X_1=x_{1,k}}(a)}{g_A(a)}\cdot f_{X_n\left|X_{n-1}=x_{n-1},\cdots,X_2=\frac{a}{x_{1,k}},X_1=x_{1,k}\right.}(x_n) \tag{A.28}$$

Equations A.27 and A.28 define the CDC associated with RVC $A_1X_3\cdots X_n$.

A.4.3 Numerical Examples

Following the examples in Section A.3.3, we consider three normal random variables *X*~*N*(12, 0.4), *Y*~*N*(3, 0.2) and *Z*~*N*(2, 0.4). The distribution of $B = XYZ$ is to be built through one surrogate $A = XY$. At first stage, *X* is discretised and the approximate probability distribution $g_A(a)$ is built by Equation A.17. The *TPB* is set as 1e-5, 1e-6, 1e-8, and 1e-10 respectively. As in Case 2 in Example 2, the borders $x_{B,k}$ are delimited by equally dividing the inverse curve and the representative x_k for the interval $\lfloor x_{B,k-1}, x_{B,k}\rfloor$ is set by

$$F_X(x_k)-F_X(x_{B,k-1})=F_X(x_{B,k})-F_X(x_k)=m_{X,k}/2$$

where $m_{X,k}$ stands for the mess probability assigned for x_k as defined in Equation A.9.

The truncated probability of *X* is attributed to the two extra points $x_{B,0}/2$ and $2x_{B,K_X}$, where K_X standard for the discretisation number of *X*.

By the above settings, the approximate probability distribution $g_A(a)$ is built at first. Based on $g_A(a)$, A is discretised in the same way as for X. The approximate probability distribution of B i.e. $g_B(b)$ is then built.

The examples are implemented by using the software package MATLAB. The simulation samples of X, Y and Z are generated directly by using MATLAB function *mvnrnd()*.

A.4.3.1 Example 4: Uncorrelated *X*, *Y* and *Z*

In this example, we assume that X, Y and Z are uncorrelated. This example aims to demonstrate the computation errors of the approximate probability distribution $g_B(b)$. For such a case, the analytical distribution $f_{B=XYZ}(b)$ has not been found in the literature. The analytical mean and the standard deviation of B however can be calculated by

$$E(A) = E(XY) = E(X)E(Y) = 36$$

$$E(B) = E(AZ) = E(A)E(Z) = 72$$

$$\mathrm{var}(A) = \mathrm{var}(XY) = \mathrm{var}(X)\cdot E(Y)^2 + \mathrm{var}(Y)\cdot E(X)^2 + \mathrm{var}(X)\cdot \mathrm{var}(Y)$$

$$\mathrm{var}(B) = \mathrm{var}(AZ) = \mathrm{var}(A)\cdot E(Z)^2 + \mathrm{var}(Z)\cdot E(A)^2 + \mathrm{var}(A)\cdot \mathrm{var}(Z)$$

$$\sigma(B) = \sqrt{\mathrm{var}(B)} = 15.4057983889184$$

Based on the analytical solution of $E(B)$ and $\sigma(B)$, we define the relative errors as

$$\mathrm{err}_{\mathrm{B,E}} = \left(E(B)\big|_{g_B(b)} - E(B)\right)\Big/E(B)$$

$$\mathrm{err}_{B,\sigma} = \left(\sigma(B)\big|_{g_B(b)} - \sigma(B)\right)\Big/\sigma(B)$$

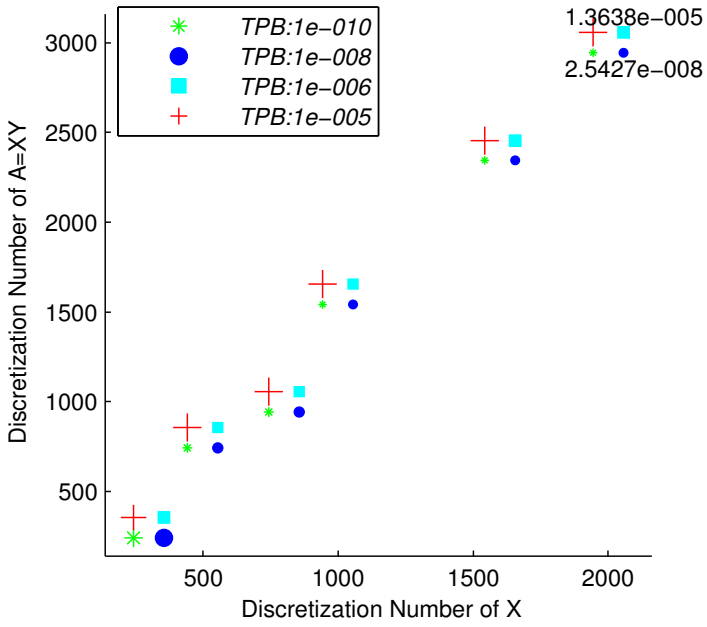
where $E(B)\big|_{g_B(b)}$ and $\sigma(B)\big|_{g_B(b)}$ stand for the mean and standard deviation of *B* that are calculated from the approximate distribution $g_B(b)$.

The relative errors are used to measure the performance of the approximate distribution of *B* corresponding to the different settings of *TPB* and the different discretisation numbers of *X* and *A* as summarized in Table A.3. The relative errors are linearly represented with the size of the marks in Fig. A.8. The minimum relative errors are 2.5e-8 and 1.2e-7 for the mean and standard deviation respectively showing a very high accuracy of the approximate probability distribution $g_B(b)$.

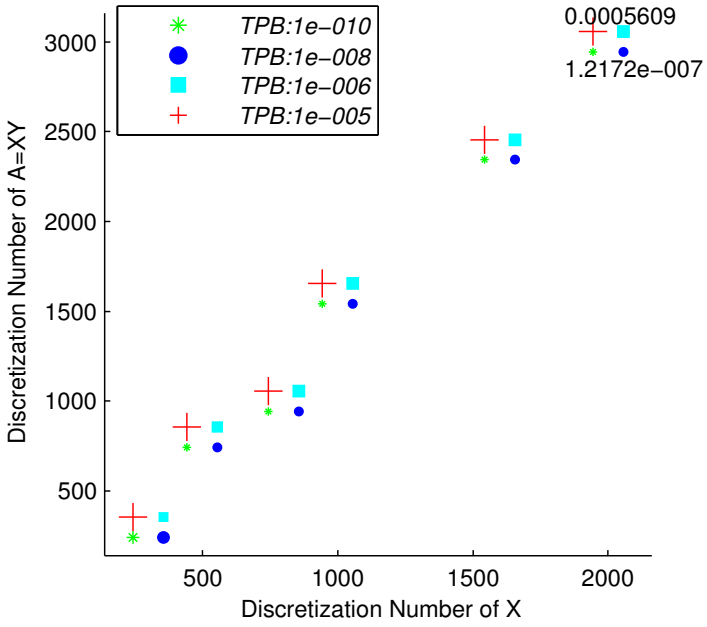
Fig. A.9 illustrates the approximate probability distribution $g_B(b)$ that is built with 2000 and 3000 discretisation points of *X* and *A* respectively. The *TPB* is set as 1e-10. The Q-Q plot of $g_B(b)$ against 60000 simulation of *B* is also drawn on Fig. A.9. We can see that the Q-Q plot coincides with the Q=Q line very well. It shows that $g_B(b)$ is a good approximation to the real distribution $f_B(b)$.

Table A.3 The discretisation numbers of *X* and *A*

No.	Discretisation Number of *X*	Discretisation Number of *A*
1	300	300
2	500	800
3	800	1000
4	1000	1600
5	1600	2400
6	2000	3000



(a) Relative error of the mean



(b) Relative error of the standard deviation

Fig. A.8. These diagrams illustrate the relative errors of $g_B(b)$ corresponding to the different settings of *TPB* and the discretisation numbers. The mark size linearly

represents the relative errors. The minimum relative errors of the mean and standard deviation are 2.5e-8 and 1.2e-7 respectively.

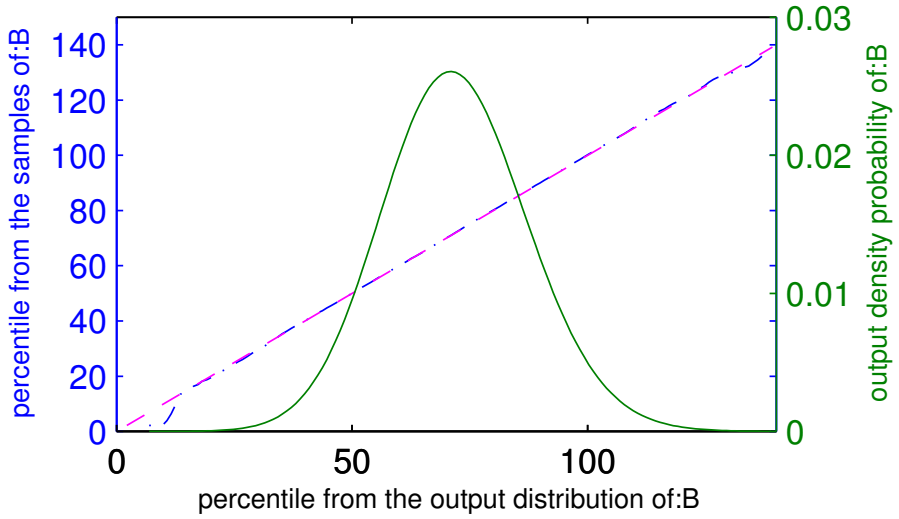


Fig. A.9. This diagram illustrates the approximate probability distribution of *B=XYZ,* where *X, Y and Z* are uncorrelated normal random variable. The green solid curve represents the density of *B*. The blue dash-dot curve represents the Q-Q plot against the 60000 simulations. The magenta dash line stands for Q=Q line.

A.4.3.2 Example 5: Correlated *X*, *Y* and *Z*

This example aims to test the performance of the approximate probability distribution of *B* when *X*, *Y* and *Z* are correlated. Correlations are set as $r(X,Y)=0.75$, $r(X,Z)=0.8$ and $r(Y,Z)=0.7$. The same settings are used as for that illustrated in Fig. A.9 in Example 4.

Fig. A.10 illustrates the outcome $g_B(b)$ and the Q-Q plot of $g_B(b)$ against 60000 simulation of *B*. We can see that the Q-Q plot coincides with the Q=Q line very well. It shows that $g_B(b)$ is a good approximation to the real distribution $f_B(b)$. Due to the correlations, the distribution curve in Fig. A.10 skews to the right further than that in

Fig. A.9 does. The mode of the distribution curve moves leftwards from around 71.2 in Fig. A.9 to around 68.2 in Fig. A.10.

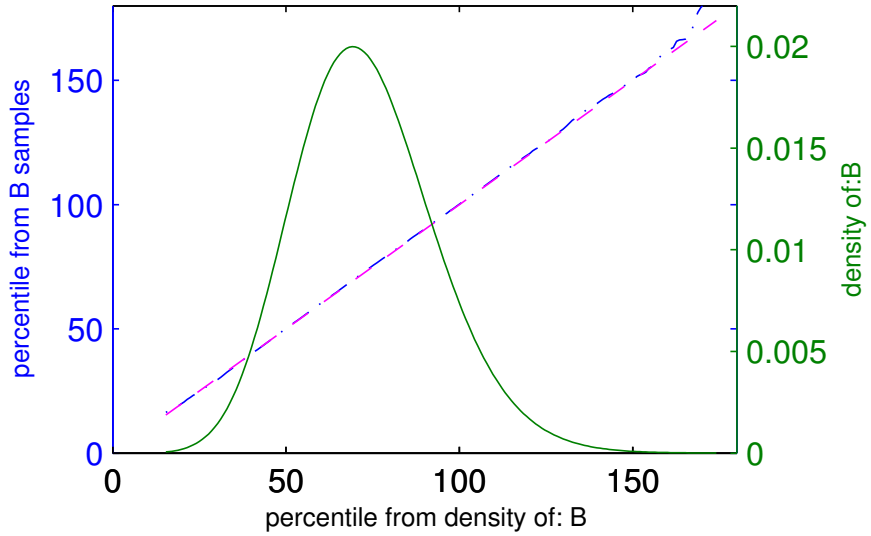


Fig. A.10. This diagram illustrates the approximate probability distribution of $B=XYZ$, when X, Y and Z are correlated normal variables. The green solid curve represents the approximate probability distribution of B. The blue dash-dot curve represents the Q-Q plot against the 60000 simulations. The magenta dash line stands for Q=Q line.

A.5 Approximate Probability Distribution of Sum of Products of Random Variables

A.5.1 Approximate Probability Distribution of Sum of Two Random Variables

Suppose

$$S = X + Y \tag{A.29}$$

where X and Y are random variable.

We suppose that X and Y have a general joint distribution defined as

$$f_{X,Y}(x,y) = f_X(x) \cdot f_{Y|X=x}(y) \tag{A.30}$$

Given $X = x$ the conditional distribution of S can be derived by shifting $f_{Y|X=x}(y)$ as

$$f_{S|X=x}(s) = f_{S|X=x}(s) = f_{Y|X=x}(s-x) \tag{A.31}$$

Based on the mixture model as defined in Equation A.3, we have

$$f_S(s) = \int_X f_{S|X=x}(s) f_X(x) dx = \int_X f_{Y|X=x}(s-x) f_X(x) dx \tag{A.32}$$

By discretising X as in Equation A.9, we can approximate the $f_S(s)$ defined in Equation A.32 by

$$g_S(s) = \sum_{k=1,K} m_k f_{Y|X=x_k}(s-x_k) \tag{A.33}$$

where m_k is the mess probability associated with the discrete point x_k.

Equation A.33 can be implemented in the same way as Equation A.10 for the product of two random variables.

A.5.2 Approximate Probability Distribution of the Sum of Multiple Random Variables

Suppose

$$S = X_1 + X_2 + \cdots + X_n \tag{A.34}$$

where $X_1, \cdots, X_n$ are random variables of the CDC defined in Equation A.23.

As for the product of multiple random variables, the sum of multiple random variables can be dealt with sequentially by replacing the first two random variables with their sum. Define the first surrogate random variable A as

$$A = X_1 + X_2 \tag{A.35}$$

The $g_A(a)$ can be built by Equation A.33. Based on Bayesian theorem [Bishop 1995], the conditional probability of $X_1 = x_{1,k}$ given $A = a$ is derived as

$$\Pr\left(X_1 = x_{1,k} \middle| A = a\right) = \frac{1}{g_A(a)} m_{1,k} f_{X_2|X_1=x_{1,k}}\left(a - x_{1,k}\right) \tag{A.36}$$

From Equation A.35, we have

$$\Pr\left(X_2 = a - x_{1,k} \middle| X_1 = x_{1,k}, A = a\right) = 1 \tag{A.37}$$

Based on Equations A.36 and A.37, we have

$$\Pr\left(X_1 = x_{1,k}, X_2 = a - x_{1,k} \middle| A = a\right) = \frac{1}{g_A(a)} m_{1,k} f_{X_2|X_1=x_{1,k}}\left(a - x_{1,k}\right) \tag{A.38}$$

From Equation A.38, the mixture model of the distribution of X_3 given $A = a$ can be approximated by Equation A.2 as

$$g_{X_3|A=a} = \sum_{k=1,K} \frac{m_{1,k} f_{X_2|X_1=x_{1,k}}\left(a - x_{1,k}\right)}{f_A(a)} \cdot f_{X_3|X_2=a-x_{1,k},X_1=x_{1,k},}\left(x_3\right) \tag{A.39}$$

The approximate conditional distribution $g_{X_i|A=a}$, $i=4,\cdots,n$ can be built as a mixture model in the same way as in Equation A.39.

By repeating the above process, the sum of multiple random variables can be dealt with sequentially in the same way as for the product of multiple random variables.

A.5.3 Process for Building Approximate Probability Distribution of Sum of Products of Random Variables

The sum of products of random variables is defined as

$$S=\prod_{i=1}^{m}X_i+\prod_{j=1}^{n}Y_j+\cdots+\prod_{\ell=1}^{k}w_\ell \tag{A.40}$$

where X_i , $i=1,\cdots,m$, Y_j , $j=1,\cdots,n$ and W_ℓ , $\ell=1,\cdots,k$ are different random variable.

According to Equation A.40, *S* can be defined by the RVC and the associated CDC as

$$\text{RVC: } X_1\cdots X_m Y_1\cdots Y_n\cdots W_1\cdots W_k \tag{A.41.a}$$

$$\text{CDC: } f_{X_1}(x_1)f_{X_2|X_1}(x_2)\ldots f_{Y_1|X_m,\cdots,X_1}(y_1)\ldots f_{W_k|W_{k-1},\cdots,X_1}(w_k) \tag{A.41.b}$$

Define

$$A=\prod_{i=1}^{m}X_i\ ,\ B=\prod_{j=1}^{n}Y_j\ ,\ D=\prod_{\ell=1}^{k}w_\ell$$

The approximate probability distribution of S can be built by alternatively applying the process for the product and the process for the sum of random variables as developed in the previous section. The process is defined as follows.

Step 1: substitute the first product A and obtain

$$\text{RVC: } AY_1 \cdots Y_n \cdots W_1 \cdots W_k \tag{A.42.a}$$

$$\text{CDC: } g_A(a)\, g_{Y_1|A}(y_1) \ldots g_{W_k|W_{k-1},\cdots Y_1,A}(w_k) \tag{A.42.b}$$

Step 2: substitute the second product B and obtain

$$\text{RVC: } AB \cdots W_1 \cdots W_k \tag{A.43.a}$$

$$\text{CDC: } g_A(a)\, g_{B|A}(b) \ldots g_{W_k|W_{k-1},\cdots,B,A}(w_k) \tag{A.43.b}$$

Step 3: substitute $S_1 = A + B$ and obtain

$$\text{RVC: } S_1 \cdots W_1 \cdots W_k \tag{A.44.a}$$

$$\text{CDC: } g_{S_1}(s) \cdots g_{W_1|\cdots,S_1}(w_1) \cdots g_{W_k|W_{k-1},\cdots,S_1}(w_k) \tag{A.44.b}$$

Step 4: repeat *Step 1* to *Step 3* until reaching the last sum representing S.

The Step 1 and Step 2 can be implemented with the algorithm for the product of random variables; while Step 3 can be implemented with the algorithm for the sum of random variables.

A.6 Conclusion

In this chapter, we propose an approximate analytical solution of the probability distribution of the product of two random variables of a general joint distribution. The approximate solution is defined by the mixture probability model; the discretisation method is investigated regarding the marginal distribution, the conditional distribution and the inverse curve $1/x$. Because normal random variables are used in the case studies planned for this research, we implement the approximate probability distribution of the product of two normal random variables for demonstration. For the product of two independent standard normal variables, we compare the approximate distribution with the analytical distribution. The comparison shows that the approximate analytical solution can approach the analytical distribution with high accuracy when we increase the number of the components in the approximate distribution.

By Bayesian theorem, the approximate analytical solution is further expanded for the product of multiple random variables and for the sum of the products of random variables. The random variables are deal with sequentially and only one random variable is discretised at each step. The method is therefore suitable for large models while it is very difficult if possible to cope with the all random variables simultaneously. We test the expanded approximate distribution model with the product of three correlated normal random variables. The outcome approximate distribution is compared with the simulations of the product.

As a conclusion, the approximate analytical solutions offer a potential generic way for building the distribution of the sum of products of random variables of a general joint distribution. They are also easy to implement on computers. The approximate analytical solutions are still not applicable. More efforts, for example, need to be exerted to measure and control the computation error.

Appendix B

Tables Relating to the Case Studies

B.1 Elicited Data for the Precursor Families Related to HET10

B.1.1 Precursor Family RVSTRENV

B.1.1.1 Variant Factors

This subsection is related to Block 3 in Fig. 6.2. We will elicit the variant factors for the precursor family RVSTRENV.

The precursor family RVSTRENV is defined for *"RV incorrectly on various types of LCs and struck by train due to environmental factors"*. As summarized in Table B.1, one variant factor class is elicited as *"RV drivers' ability to respond to the prevailing weather conditions"*. The level crossings are put into two categories: user worked crossings (UWC) and non-UWC [RSSB 2004]. The two categories of LCs have their own users. Accordingly the two groups of users are defined as UWC users and non-UWC users.

Table B.1 Elicited factor class for the family RVSTRENV

Family	RVSTRENV – RV incorrectly on various types of LCs and struck by train due to environmental factors
No.	**Variant Class**
2.	RV drivers' ability to respond to the prevailing weather conditions.

B.1.1.2 Correlations between the Precursors and the Associated Variant Factors

This subsection is related to Block 4a in Fig. 6.2 because one variant factor class has been elicited for the family RVSTRENV. We will elicit the correlations between the precursors and the associated variant factors.

A medium correlation is elicited between the precursors on UWC LCs and the UWC users; a medium correlation is also elicited between the precursors on non-UWC LCs and the non-UWC users. The elicited results are summarized in Table B.2. Because the elicited correlations are positive in this case study we do not mark the correlation sign explicitly.

Table B.2 Elicited correlations between the variant factors and the associated precursors of the family RVSTRENV

	Precursor Family	RVSTRENV – RV incorrectly on LC and struck by train due to environmental factors	
	Variant Class	RV drivers' ability to respond to the prevailing weather conditions.	
NO.	**Variant Factors**	**Associated precursor**	**Correlation**
3.	Concerning UWC LCs	RVSTRENV on UWC LCs	M
4.	Concerning Non-UWC LCs	RVSTRENV on Non-UWC LCs	M

B.1.1.3 Correlations within the Variant Factor Class

This subsection is related to Block 5 in Fig. 6.2. We will elicit the correlations within the factor class of *"RV drivers' ability to respond to the prevailing weather conditions"*.

A medium correlation between UWC users and non-UWC users is elicited as summarized in Table B.3.

Table B.3 The correlations within the factor class for the family RVSTRENV

Family name	RVSTRENV – RV incorrectly on LCs and struck by train due to environmental factors

Variant class	RV drivers' ability to respond to the prevailing weather conditions
Variant factor 1	On UWC LCs
Variant factor 2	On Non-UWC LCs
The correlation between the above two variant factors	M

B.1.2 Precursor Family TOVRSPD

B.1.2.1 Variant Factors

This subsection is related to Block 3 in Fig. 6.2. We will elicit the variant factors for the precursor family TOVRSPD.

The precursor family TOVRSPD is defined as *"Train over-speeding causes RV struck by train on various types of LCs"*. For this family, no significant variant factor is identified by the experts.

B.1.2.2 Proportion of the Variance of Each Precursor due to the Family Commonality

This subsection is related to Block 4b in Fig. 6.2 because no variant factor is identified for the family TOVRSPD.

As proposed previously, the proportion of the variance due to the family commonality is elicited for each family member as summarized in Table B.4.

Table B.4 Proportion of the variance due to the family commonality elicited for each precursor in the family TOVRSPD

Family	**TOVRSPD – Train over-speeding causes RV struck by train on various types of LCs**	
No.	**Precursor**	**Proportion of the variance due to the family commonality** (100% for a very strong family; 90% for a less strong family)
1.	TOVRSPD on ABCL	0.9
2.	TOVRSPD on AHB	0.9
3.	TOVRSPD on AOCL	0.9
4.	TOVRSPD on MG/B	0.9

B.1.3 Precursor Family RVSTRAN

B.1.3.1 Variant Factors

This subsection is related to Block 3 in Fig. 6.2. We will elicit the variant factors for the precursor family RVSTRAN.

The precursor family RVSTRAN is defined as *"RV stranded on LC causes RV struck by train on various types of LCs"*. For this family, two variant factor classes are identified as summarized in Table B.5 as:

- FC1: Propensity for there to be blocking back on a crossing, i.e. where you get traffic jams extending back over level crossings
- FC2: Profile of the RV drivers using the crossing, i.e. the propensity of certain drivers to violate rules

Table B.5 Factor classes of the precursor family RVSTRAN

Family	RVSTRAN – RV stranded on LC causes RV struck by train on various types of LCs
No.	**Variant Classes**
1.	Propensity for there to be blocking back on a level crossing
2.	Propensity of the drivers violating the rules when using the level crossing

B.1.3.2 Correlations between the Precursors and the Associated Variant Factors

This subsection is related to Block 4a in Fig. 6.2 because one variant factor class has been elicited for the family. We will elicit the correlations between the precursors and the associated variant factors.

For the variant factors regarding the propensity for there to be blocking back on a crossing, the correlations between the precursors and the associated variant factors are elicited and summarized in Table B.6.

For the variant factors regarding the propensity of the drivers violating the rules when using the level crossing, the UWC user group and non-UWC user group are defined as for the family RVSTRENV. The correlations between the precursors and the associated variant factors are elicited and summarized in Table B.7.

Table B.6 Elicited correlations between the precursors in the family RVSTRAN and the associated variant factors regarding the propensity for there to be blocking back on a crossing

Precursor Family		RVSTRAN – RV stranded on various LC	
Variant Factor Class		Propensity for there to be blocking back on a crossing	
NO.	**Variant Factor**	**Associated precursor**	**Correlation**
1.	Propensity for there to be blocking back on ABCL	RVSTRAN on ABCL	W (ABCL tend to quite rural and hence light use by traffic)
2.	Propensity for there to be blocking back on AHB	RVSTRAN on AHB	S (blocking back seems to happen predominantly on AHB crossings)
3.	Propensity for there to be blocking back on AOCL	RVSTRAN on AOCL	W (AOCL tend to be quite rural and hence light use by traffic)
4.	Propensity for there to be blocking back on MG/B	RVSTRAN on MG/B	VW (manual barriers would have to be closed round vehicle)
5.	Propensity for there to be blocking back on UWC	RVSTRAN on UWC	VW (likely to be single use)
6.	Propensity for there to be blocking back on UWC-MWL	RVSTRAN on UWC-MWL	VW (likely to be single use)
7.	Propensity for there to be blocking back on UWC+T	RVSTRAN on UWC+T	VW (likely to be single use)

Table B.7 Elicited correlations between the precursors in Family RVSTRAN and the associated variant factors regarding the propensity of the drivers violating the rules when using the level crossing

Precursor Family		RVSTRAN – RV stranded on various LCs	
Variant Factor Class		Propensity of the drivers violating the rules when using the level crossing	
NO.	**Variant Factor** Propensity of the drivers violating the rules when using the LC on	**Associated precursor**	**Correlation**
1.	ABCL	RVSTRAN on ABCL	M
2.	AHB	RVSTRAN on AHB	M
3.	AOCL	RVSTRAN on AOCL	VW

4.	MG/B	RVSTRAN on MG/B	VW
5.	UWC	RVSTRAN on UWC	S
6.	UWC-MWL	RVSTRAN on UWC-MWL	S
7.	UWC+T	RVSTRAN on UWC+T	S

B.1.3.3 Correlations within the Variant Factor Classes

This subsection is related to Block 5 in Fig. 6.2. We will elicit the correlations within the variant factor classes.

For the variant factor class regarding the propensity for there to be blocking back on a crossing, the correlations within the factor class are elicited and summarized in Table B.8.

For the variant factor class regarding the propensity of the drivers violating the rules when using the level crossing, the correlations within the class are elicited and summarized in Table B.9.

Table B.8 Correlations within the variant factor class regarding the propensity for there to be blocking back on a crossing

Family	RVSTRAN – RV stranded on various LC					
Variant class	Propensity for there to be blocking back on a crossing					
Variant factors	On AHB	On AOCL	On MG/B	On UWC	On UWC-MWL	On UWC+T
On ABCL	W	M (due to similar location of LCs)	VW	W	W	W
On AHB		VW	VW	W	W	W
On AOCL			VW	VW	VW	VW
On MG/B				VW	VW	VW
On UWC					VS (due to similar UWC usage)	VS (due to similar UWC usage)
On UWC-MWL						VS (due to similar UWC usage)

Table B.9 Correlations within the variant factor class regarding the propensity of the drivers violating the rules when using the level crossing

Family	RVSTRAN – RV stranded on various LC
Variant class	Propensity of the drivers violating the rules when using the level crossing
Variant factors	On Non-UWCs
On UWCs	M

B.1.4 Precursor Family SPADPROT

B.1.4.1 Variant Factors

This subsection is related to Block 3 in Fig. 6.2. We will elicit the variant factors for the precursor family SPADPROT.

The precursor family SPADPROT is defined as *"SPAD at signal protecting the LC causes RV struck by train on various types of LCs"*. As summarized in Table B.10, one factor class is identified by the experts as *"the effectiveness of the signal protecting the LCs"*.

Table B.10 Elicited factor class of the family SPADPROT

Family	**SPADPROT – SPAD at signal protecting the LC**
No.	**Variant Factor Class**
1.	Effectiveness of signals protecting the LC

The Correlations between the Precursors and the Associated Variant Factors
This subsection is related to Block 4a in Fig. 6.2 because one variant factor class has been elicited for the family. We will elicit the correlations between the precursors and the associated variant factors.

The correlations between the precursors and the associated variant factors are elicited and summarized in Table B.11.

Table B.11 Elicited correlations between the precursors in Family SPADPROT and the associated variant factors

Precursor Family		SPADPROT – SPAD at signal protecting the LC	
Variant Class		**Effectiveness of signals protecting the LC**	
NO.	**Variant Factor**	**Associated precursor**	**Correlation**
1.	Effectiveness of the signals protecting ABCL	SPADPROT on ABCL	VW
2.	Effectiveness of the signals protecting AOCL	SPADPROT on AOCL	VW
3.	Effectiveness of the signals protecting MG/B	SPADPROT on MG/B	VW

B.1.4.2 Correlations within the Variant Factor Class

This subsection is related to Block 5 in Fig. 6.2. We will elicit the correlations within the factor class of the effectiveness of the signals protecting LCs.

The correlations within the factor class are elicited and summarized in Table B.12.

Table B.12 Elicited correlations within the variant factor class regarding the effectiveness of the signals protecting LCs

Family name	SPADPROT – SPAD at signal protecting the LC		
Variant class	Quality of the drivers using the LC (propensity of certain drivers to violate rules)		
Variant factors	On AOCL	On MG/B	Note: The physical similarity of the signals protecting the different types of level crossings is a good clue on the correlation.
On ABCL	VW	VW	
On AOCL		VW	

B.1.5 Precursor Family SIGERR

B.1.5.1 Variant Factors

This subsection is related to Block 3 in Fig. 6.2. We will elicit the variant factors for the precursor family SIGERR.

The precursor family SIGERR is defined as *"Signalman or Crossing keeper error causes RV struck by train on various types of LCs"*. As summarized in Table B.13, one factor class is identified as *"Complexity of operating the crossing"*.

Table B.13 Elicited factor class of the family SIGERR

Family	**SIGERR – Signalman or Crossing keeper error causes RV struck by train on various types of LCs**
No.	**Variant Factor Class**
1.	complexity of operating the crossing

The Correlations between the Precursors and the Associated Variant Factors
This subsection is related to Block 4a in Fig. 6.2 because one variant factor class has been elicited for the family. We will elicit the correlations between the precursors and the associated variant factors.

The correlations between the precursors and the associated variant factors are elicited and summarized in Table B.14.

Table B.14 Elicited correlations between the precursors in Family SIGERR and the associated variant factors

	Precursor Family	**SIGERR – Signalman or Crossing keeper error causes RV struck by train on various types of crossing**	
	Variant Factor Calss	**complexity of operating the crossings**	
NO.	**Variant Factor**	**Associated precursor**	**Correlation**
1.	complexity of operating the crossing on AHB	SPADPROT on AHB	M
2.	complexity of operating the crossing on MGB	SPADPROT on MG/B	M

3.	complexity of operating the crossing on UWC+T	SPADPROT on UWC+T	M

B.1.5.2 Correlations within the Variant Factor Class

This subsection is related to Block 5 in Fig. 6.2. We will elicit the correlations within the factor class of the complexity of operating the crossings.

The correlations within the factor class are elicited and summarized in Table B.15.

Table B.15 Elicited correlations within the factor class regarding the complexity of operating the crossings

Family name	SIGERR – Signalman or Crossing keeper error causes RV struck by train on various types of crossing		
Variant class	complexity of operating the crossings		
Variant factors	On MG/B	On UWC+T	Note: the similarity of operating level crossing may be a good clue on the correlation.
On AHB	VW (different operation)	S (both remotely operated)	
On MG/B		VW (different operation)	

B.1.6 Precursor Family RVDRVERR and RVDRVDEL

The experts believe that the precursors *"RV Driver error causing RV struck by train on LCs"* and the precursors *"RV driver deliberate action on various types of crossing"* are identical on the same level crossing in terms of the uncertainty modelling. Correspondingly these two groups of precursors are defined within in one family RVDRVERR and RVDRVDEL.

B.1.6.1 Variant Factors

This subsection is related to Block 3 in Fig. 6.2. We will elicit the variant factors for the precursor family RVDRVERR and RVDRVDEL.

As summarized in Table B.16, one factor class is identified by the experts as *"the quality of the users"*. The UWC user group and non-UWC user group are defined as for the families RVSTRENV and RVSTRAN.

Table B.16 Elicited factor class of the family RVDRVERR & RVDRVDEL

Family	DRERRSTR & RVDRVDEL: RV Driver error causing RV struck by train on LCs and RV driver deliberate action on various types of crossing
No.	**Variant Factor Class**
1.	Quality of the drivers using the crossing (propensity of certain drivers to violate rules)

The Correlations between the Precursors and the Associated Variant Factors
This subsection is related to Block 4a in Fig. 6.2 because one variant factor class has been elicited for the family. We will elicit the correlations between the precursors and the associated variant factors.

The correlations between the precursors and the associated variant factors are elicited and summarized in Table B.17.

Table B.17 Elicited correlations between the precursors in Family RVDRVERR and RVDRVDEL and the associated variant factors

	Precursor Family	DRERRSTR & RVDRVDEL	
	Variant Factor Class	Quality of the drivers using the crossing (propensity of certain drivers to violate rules)	
NO.	**Variant Factor**	**Associated precursor**	**Correlation**
1.	Quality of the RV drivers using UWC level crossings	DRERRSTR & RVDRVDEL on UWC level crossings	Very Strong
2.	Quality of the RV drivers using Non-UWC level crossings	DRERRSTR & RVDRVDEL on Non-UWC level crossings	Very Strong

B.1.6.2 Correlations within the Variant Factor Class

This subsection is related to Block 5 in Fig. 6.2. We will elicit the correlations within the factor class of

The correlations within the factor class are elicited and summarized in Table B.18.

Table B.18 Elicited correlations within the factor class regarding the quality of the users of the LCs

Family name	DRERRSTR & RVDRVDEL
Variant factor class	Quality of the drivers using the crossing (propensity of certain drivers to violate rules)
Variant factors 1	On UWC level crossings
Variant factors 2	On Non-UWC level crossings
The correlation between the above two variant factors (Tip: think of the similarity between the two groups of drivers)	M – the reason for this is that the users using the UWC crossings will also form part of the driver population who use non-UWC crossings. Therefore information regarding the uncertainty of the UWC error rate is likely to inform us of the uncertainty in the Non-UWC error rate.

B.1.7 Precursor Family RVDRVSUI

B.1.7.1 Variant Factors

This subsection is related to Block 3 in Fig. 6.2. We will elicit the variant factors for the precursor family RVDRVSUI.

The precursor family RVDRVSUI is defined as *"RV driver suicide causes RV struck by train on various types of LCs"*. For this family, no significant variant class is identified by the experts.

The Proportion of the Variance of Each Precursor due to the Family Commonality
This subsection is related to Block 4b in Fig. 6.2 because no variant factor is identified for the family.

As proposed previously, the proportion of the variance due to the family commonality is elicited for each precursor as summarized in Table B.19.

Table B.19 Elicited proportion of the variance due to the family commonality for the family RVDRVSUI

Family Name	**RVDRVSUI - RV driver suicide on various types of LC**	
No.	**Precursor**	**Proportion of covariance due to the family**

		commonality(Number between 0 and 1)
1.	RVDRVSUI on ABCL	0.95
2.	RVDRVSUI on AHB	0.95
3.	RVDRVSUI on AOCL	0.95
4.	RVDRVSUI on O	0.95
5.	RVDRVSUI on UWC	0.95
6.	RVDRVSUI on UWC-MWL	0.95
7.	RVDRVSUI on UWC-T	0.95

B.1.8 Precursor Family LTBRFAI

B.1.8.1 Variant Factors

This subsection is related to Block 3 in Fig. 6.2. We will elicit the variant factors for the precursor family LTBRRAI.

The precursor family LTBRFAI is defined as *"That Light/barriers fail to operate causes RV struck by train on various types of LCs"*. As summarized in Table B.20, one factor class is identified by the experts as *"the technical workings of the lights and barriers and their failure rates"*.

Table B.20 Elicited factor class of the family LTBRFAI

Family	LTBRFAI – Light/barriers fail to operate on various types of LC
No.	**Variant Class**
1.	Technical workings of the various lights and barriers and their likely failure rates

The Correlations between the Precursors and the Associated Variant Factors
This subsection is related to Block 4a in Fig. 6.2 because one variant factor class has been elicited for the family. We will elicit the correlations between the precursors and the associated variant factors.

The correlations between the precursors and the associated variant factors are elicited as summarized in Table B.21.

Table B.21 Elicited correlations between the precursors in Family LTBRFAI and the associated variant factors

Precursor Family		LTBRFAI – Light/barriers fail to operate on various types of LCs	
Variant Factor Class		Technical workings of the various lights and barriers and their likely failure rates	
NO.	**Variant Factor**	**Associated precursor**	**Correlation**
3.	Technical workings and likely failure rates on ABCL	LTBRFAI on ABCL	M
4.	Technical workings and likely failure rates on AHB	LTBRFAI on AHB	M
5.	Technical workings and likely failure rates on AOCL	LTBRFAI on AOCL	M
6.	Technical workings and likely failure rates on MG/B	LTBRFAI on MG/B	W (as the crossing is under manual control

B.1.8.2 Correlations within the Variant Factor Class

This subsection is related to Block 5 in Fig. 6.2. We will elicit the correlations within the factor class of the technical workings and likely failure rates between the LCs.

The correlations within the factor class are elicited as summarized in Table B.22.

Table B.22 Elicited correlations within the factor class regarding the technical workings and likely failure rates between the LCs

Family	LTBRFAI – Light/barriers fail to operate on various types of LC			
Variant class	Technical workings of the various lights and barriers and their likely failure rates			
Variant factors	On AHB	On AOCL	On MG/B	The technical similarity of the equipment of light and barrier between the level crossing types is a good clue to assess the correlation.
On ABCL	VS (similar lights & barrier)	W (similar lights – barriers are different)	M (similar lights)	
On AHB		W (similar lights – barriers are different)	M (similar lights)	
On AOCL			M (similar lights)	

B.2 Rule Sets Related to HET10

Table B.23 Rule sets related to HET10

Rule Set Code	Description
T10-LCPAU-2	Train strikes normal road vehicle (car/HGV) on AOCL or ABCL crossing, with fire
T10-LCPAU-3	Train strikes large road vehicle above buffer height (low loader, JCB, etc.) on an AOCL or ABCL crossing, no fire
T10-LCPAU-4	Train strikes large road vehicle above buffer height (low loader, JCB, etc.) on an AOCL or ABCL crossing, with fire
T10-LCPAU-5	Train strikes HGV carrying toxic hazardous goods on AOCL or ABCL crossing, no fire
T10-LCPAU-6	Train strikes HGV carrying toxic hazardous goods on AOCL or ABCL crossing, with fire
T10-LCPAU-7	Train strikes HGV carrying flammable hazardous goods on AOCL or ABCL crossing, with fire
T10-LCPRO-2	Train strikes normal road vehicle (car/HGV) on AHB, FP, MB/MCB/CCTV & all UWC (incl. MWL + T & UWC), with fire
T10-LCPRO-3	Train strikes large road vehicle above buffer height (low loader, JCB, etc.) on AHB, FP, MB/MCB/CCTV & all UWC (incl. MWL + T & UWC), no fire
T10-LCPRO-4	Train strikes large road vehicle above buffer height (low loader, JCB, etc.) on AHB, FP, MB/MCB/CCTV & all UWC (incl. MWL + T & UWC), with fire
T10-LCPRO-5	Train strikes HGV carrying toxic hazardous goods on AHB, FP, MB/MCB/CCTV & all UWC (incl. MWL + T & UWC), no fire
T10-LCPRO-6	Train strikes HGV carrying toxic hazardous goods on AHB, FP, MB/MCB/CCTV & all UWC (incl. MWL + T & UWC), with fire
T10-LCPRO-7	Train strikes HGV carrying flammable hazardous goods on AHB, FP, MB/MCB/CCTV & all UWC (incl. MWL + T & UWC) crossing, with fire

B.3 Rule Sets Related to HET12

Table B.24 Rule sets related to HET12

Code	Description
T12-BG-1	Passenger train derailment on rail bridge
T12-OF-01	PT Fast derailment on open track inside train clearances
T12-OF-02	PT Fast derailment on open track outside train clearances, towards cess side, carriage not on its side
T12-OF-03	PT Fast derailment on open track outside train clearances, towards cess side, carriage not on its side, hits line side structure, no fire
T12-OF-04	PT Fast derailment on open track outside train clearances, towards cess side, carriage not on its side, hits line side structure, with fire
T12-OF-05	PT Fast derailment on open track outside train clearances, towards cess side, carriage not on its side, hits line side structure, structure collapses onto train, no fire
T12-OF-06	PT Fast derailment on open track outside train clearances, towards cess side, carriage not on its side, hits line side structure, structure collapses onto train, with fire
T12-OF-07	PT Fast derailment on open track outside train clearances, towards cess side, carriage on its side, no fire
T12-OF-08	PT Fast derailment on open track outside train clearances, towards cess side, carriage on its side, with fire
T12-OF-09	PT Fast derailment on open track outside train clearances, towards cess side, carriage on its side, hits line side structure, no fire
T12-OF-10	PT Fast derailment on open track outside train clearances, towards cess side, carriage on its side, hits line side structure, with fire
T12-OF-11	PT Fast derailment on open track outside train clearances, towards cess side, carriage on its side, hits line side structure, structure collapses onto train, no fire
T12-OF-12	PT Fast derailment on open track outside train clearances, towards cess side, carriage on its side, hits line side structure, structure collapses onto train, with fire
T12-OF-13	PT Fast derailment on open track outside train clearances, towards adjacent line, carriage not on its side, secondary collision with PT, no fire
T12-OF-14	PT Fast derailment on open track outside train clearances, towards adjacent line, carriage not on its side, secondary collision with PT, with fire
T12-OF-15	PT Fast derailment on open track outside train clearances, towards adjacent

Code	Description
	line, carriage not on its side, secondary collision with FT carrying non-hazardous goods, no fire
T12-OF-16	PT Fast derailment on open track outside train clearances, towards adjacent line, carriage not on its side, secondary collision with FT carrying non-hazardous goods, with fire
T12-OF-17	PT Fast derailment on open track outside train clearances, towards adjacent line, carriage not on its side, secondary collision with FT carrying toxic haz goods, no fire
T12-OF-18	PT Fast derailment on open track outside train clearances, towards adjacent line, carriage not on its side, secondary collision with FT carrying toxic haz goods, with fire
T12-OF-19	PT Fast derailment on open track outside train clearances, towards adjacent line, carriage on side, secondary collision with FT carrying flammable haz goods, no fire
T12-OF-20	PT Fast derailment on open track outside train clearances, towards adjacent line, carriage not on its side, secondary collision with FT carrying flammable haz goods, with fire
T12-OF-21	PT Fast derailment on open track outside train clearances, towards adjacent line, carriage on its side, secondary collision with PT, no fire
T12-OF-22	PT Fast derailment on open track outside train clearances, towards adjacent line, carriage on its side, secondary collision with PT, with fire
T12-OF-23	PT Fast derailment on open track outside train clearances, towards adjacent line, carriage on its side, secondary collision with FT carrying non-hazardous goods, no fire
T12-OF-24	PT Fast derailment on open track outside train clearances, towards adjacent line, carriage on its side, secondary collision with FT carrying non-hazardous goods, with fire
T12-OFAUTO-01	PT Fast derailment on open track inside train clearances
T12-OFAUTO-02	PT Fast derailment on open track outside train clearances, towards cess side, carriage not on its side
T12-OFAUTO-03	PT Fast derailment on open track outside train clearances, towards cess side, carriage not on its side, hits line side structure, no fire
T12-OFAUTO-04	PT Fast derailment on open track outside train clearances, towards cess side, carriage not on its side, hits line side structure, with fire

Code	Description
T12-OFAUTO-05	PT Fast derailment on open track outside train clearances, towards cess side, carriage not on its side, hits line side structure, structure collapses onto train, no fire
T12-OFAUTO-06	PT Fast derailment on open track outside train clearances, towards cess side, carriage not on its side, hits line side structure, structure collapses onto train, with fire
T12-OFAUTO-07	PT Fast derailment on open track outside train clearances, towards cess side, carriage on its side, no fire
T12-OFAUTO-08	PT Fast derailment on open track outside train clearances, towards cess side, carriage on its side, with fire
T12-OFAUTO-09	PT Fast derailment on open track outside train clearances, towards cess side, carriage on its side, hits line side structure, no fire
T12-OFAUTO-10	PT Fast derailment on open track outside train clearances, towards cess side, carriage on its side, hits line side structure, with fire
T12-OFAUTO-11	PT Fast derailment on open track outside train clearances, towards cess side, carriage on its side, hits line side structure, structure collapses onto train, no fire
T12-OFAUTO-12	PT Fast derailment on open track outside train clearances, towards cess side, carriage on its side, hits line side structure, structure collapses onto train, with fire
T12-OFAUTO-13	PT Fast derailment on open track outside train clearances, towards adjacent line, carriage not on its side, secondary collision with PT, no fire
T12-OFAUTO-14	PT Fast derailment on open track outside train clearances, towards adjacent line, carriage not on its side, secondary collision with PT, with fire
T12-OFAUTO-15	PT Fast derailment on open track outside train clearances, towards adjacent line, carriage not on its side, secondary collision with FT carrying non-hazardous goods, no fire
T12-OFAUTO-16	PT Fast derailment on open track outside train clearances, towards adjacent line, carriage not on its side, secondary collision with FT carrying non-hazardous goods, with fire
T12-OFAUTO-17	PT Fast derailment on open track outside train clearances, towards adjacent line, carriage not on its side, secondary collision with FT carrying toxic haz goods, no fire
T12-OFAUTO-18	PT Fast derailment on open track outside train clearances, towards adjacent line, carriage not on its side, secondary collision with FT carrying toxic haz

Code	Description
	goods, with fire
T12-OFAUTO-19	PT Fast derailment on open track outside train clearances, towards adjacent line, carriage on side, secondary collision with FT carrying flammable haz goods, no fire
T12-OFAUTO-20	PT Fast derailment on open track outside train clearances, towards adjacent line, carriage not on its side, secondary collision with FT carrying flammable hazardous goods, with fire
T12-OFAUTO-21	PT Fast derailment on open track outside train clearances, towards adjacent line, carriage on its side, secondary collision with PT, no fire
T12-OFAUTO-22	PT Fast derailment on open track outside train clearances, towards adjacent line, carriage on its side, secondary collision with PT, with fire
T12-OFAUTO-23	PT Fast derailment on open track outside train clearances, towards adjacent line, carriage on its side, secondary collision with FT carrying non-hazardous goods, no fire
T12-OFAUTO-24	PT Fast derailment on open track outside train clearances, towards adjacent line, carriage on its side, secondary collision with FT carrying non-hazardous goods, with fire
T12-OS-03	PT Slow derailment on open track outside train clearances, towards cess side, carriage not on its side, hits line side structure, no fire
T12-OS-04	PT Slow derailment on open track outside train clearances, towards cess side, carriage not on its side, hits line side structure, with fire
T12-OS-05	PT Slow derailment on open track outside train clearances, towards cess side, carriage not on its side, hits line side structure, structure collapses onto train, no fire
T12-OS-06	PT Slow derailment on open track outside train clearances, towards cess side, carriage not on its side, hits line side structure, structure collapses onto train, with fire
T12-OS-07	PT Slow derailment on open track outside train clearances, towards cess side, carriage on its side, no fire
T12-OS-08	PT Slow derailment on open track outside train clearances, towards cess side, carriage on its side, with fire
T12-OS-09	PT Slow derailment on open track outside train clearances, towards cess side, carriage on its side, hits line side structure, no fire
T12-OS-10	PT Slow derailment on open track outside train clearances, towards cess

Code	Description
	side, carriage on its side, hits line side structure, with fire
T12-OS-11	PT Slow derailment on open track outside train clearances, towards cess side, carriage on its side, hits line side structure, structure collapses onto train, no fire
T12-OS-12	PT Slow derailment on open track outside train clearances, towards cess side, carriage on its side, hits line side structure, structure collapses onto train, with fire
T12-OS-13	PT Slow derailment on open track outside train clearances, towards adjacent line, carriage not on its side, secondary collision with PT, no fire
T12-OS-14	PT Slow derailment on open track outside train clearances, towards adjacent line, carriage not on its side, secondary collision with PT, with fire
T12-OS-15	PT Slow derailment on open track outside train clearances, towards adjacent line, carriage not on its side, secondary collision with FT carrying non-hazardous goods, no fire
T12-OS-16	PT Slow derailment on open track outside train clearances, towards adjacent line, carriage not on its side, secondary collision with FT carrying non-hazardous goods, with fire
T12-OS-17	PT Slow derailment on open track outside train clearances, towards adjacent line, carriage not on its side, secondary collision with FT carrying toxic haz goods, no fire
T12-OS-18	PT Slow derailment on open track outside train clearances, towards adjacent line, carriage not on its side, secondary collision with FT carrying toxic haz goods, with fire
T12-OS-19	PT Slow derailment on open track outside train clearances, towards adjacent line, carriage on its side, secondary collision with FT carrying flammable haz goods, no fire
T12-OS-20	PT Slow derailment on open track outside train clearances, towards adjacent line, carriage not on its side, secondary collision with FT carrying flammable haz goods, with fire
T12-OS-21	PT Slow derailment on open track outside train clearances, towards adjacent line, carriage on its side, secondary collision with PT, no fire
T12-OS-22	PT Slow derailment on open track outside train clearances, towards adjacent line, carriage on its side, secondary collision with PT, with fire
T12-OS-23	PT Slow derailment on open track outside train clearances, towards adjacent line, carriage on its side, secondary collision with FT carrying non-hazardous

Code	Description
	goods, no fire
T12-OS-24	PT Slow derailment on open track outside train clearances, towards adjacent line, carriage on its side, secondary collision with FT carrying non-hazardous goods, with fire
T12-ST-1	PT Fast derailment in single track tunnel inside train clearances
T12-ST-2	PT Fast derailment in single track tunnel outside train clearances, carriage in contact with tunnel wall, no fire
T12-ST-3	PT Fast derailment in single track tunnel outside train clearances, carriage in contact with tunnel wall, with fire
T12-TT-1	PT Fast derailment in twin track tunnel inside train clearances
T12-TT-2	PT Fast derailment in twin track tunnel outside train clearances, towards cess side, carriage in contact with tunnel wall, no fire
T12-TT-3	PT Fast derailment in twin track tunnel outside train clearances, towards cess side, carriage in contact with tunnel wall, with fire
T12-TT-7	PT Fast derailment in twin track tunnel outside train clearances, towards cess side, carriage on its side, no fire
T12-TT-8	PT Fast derailment in twin track tunnel outside train clearances, towards cess side, carriage on its side, with fire
T12-TT-13	PT Fast derailment in twin track tunnel outside train clearances, towards adjacent line, carriage not on its side, secondary collision with PT, no fire
T12-TT-14	PT Fast derailment in twin track tunnel outside train clearances, towards adjacent line, carriage not on its side, secondary collision with PT, with fire
T12-TT-15	PT Fast derailment in twin track tunnel outside train clearances, towards adjacent line, carriage not on its side, secondary collision with FT carrying non-hazardous goods, no fire
T12-TT-16	PT Fast derailment in twin track tunnel outside train clearances, towards adjacent line, carriage not on its side, secondary collision with FT carrying non-hazardous goods, with fire
T12-TT-17	PT Fast derailment in twin track tunnel outside train clearances, towards adjacent line, carriage not on its side, secondary collision with FT carrying toxic hazardous goods, no fire
T12-TT-18	PT Fast derailment in twin track tunnel outside train clearances, towards adjacent line, carriage not on its side, secondary collision with FT carrying toxic hazardous goods, with fire

Code	Description
T12-TT-19	PT Fast derailment in twin track tunnel outside train clearances, towards adjacent line, carriage on its side, secondary collision with FT carrying flammable hazardous goods, no fire
T12-TT-20	PT Fast derailment in twin track tunnel outside train clearances, towards adjacent line, carriage not on its side, secondary collision with FT carrying flammable hazardous goods, with fire
T12-TT-21	PT Fast derailment in twin track tunnel outside train clearances, towards adjacent line, carriage on its side, secondary collision with PT, no fire
T12-TT-22	PT Fast derailment in twin track tunnel outside train clearances, towards adjacent line, carriage on its side, secondary collision with PT, with fire
T12-TT-23	PT Fast derailment in twin track tunnel outside train clearances, towards adjacent line, carriage on its side, secondary collision with FT carrying non-hazardous goods, no fire
T12-TT-24	PT Fast derailment in twin track tunnel outside train clearances, towards adjacent line, carriage on its side, secondary collision with FT carrying non-hazardous goods, with fire
T14-SF-1	PT Fast derailment at station inside train clearances
T14-SF-2	PT Fast derailment at station outside train clearances, towards cess side, carriage not on its side
T14-SF-3	PT Fast derailment at station outside train clearances, towards cess side, carriage not on its side, hits line side structure, no fire
T14-SF-4	PT Fast derailment at station outside train clearances, towards cess side, carriage not on its side, hits line side structure, with fire
T14-SF-7	PT Fast derailment at station outside train clearances, towards cess side, carriage on its side, no fire
T14-SF-8	PT Fast derailment at station outside train clearances, towards cess side, carriage on its side, with fire
T14-SF-9	PT Fast derailment at station outside train clearances, towards cess side, carriage on its side, hits line side structure, no fire
T14-SF-10	PT Fast derailment at station outside train clearances, towards cess side, carriage on its side, hits line side structure, with fire
T14-SF-13	PT Fast derailment at station outside train clearances, towards adjacent line, carriage not on its side, secondary collision with PT, no fire
T14-SF-14	PT Fast derailment at station outside train clearances, towards adjacent line,

Code	Description
	carriage not on its side, secondary collision with PT, with fire
T14-SF-15	PT Fast derailment at station outside train clearances, towards adjacent line, carriage not on its side, secondary collision with FT carrying non-hazardous goods, no fire
T14-SF-16	PT Fast derailment at station outside train clearances, towards adjacent line, carriage not on its side, secondary collision with FT carrying non-hazardous goods, with fire
T14-SF-17	PT Fast derailment at station outside train clearances, towards adjacent line, carriage not on its side, secondary collision with FT carrying toxic hazardous goods, no fire
T14-SF-18	PT Fast derailment at station outside train clearances, towards adjacent line, carriage not on its side, secondary collision with FT carrying toxic hazardous goods, with fire
T14-SF-19	PT Fast derailment at station outside train clearances, towards adjacent line, carriage on its side, secondary collision with FT carrying flammable hazardous goods, no fire
T14-SF-20	PT Fast derailment at station outside train clearances, towards adjacent line, carriage not on its side, secondary collision with FT carrying flammable hazardous goods, with fire
T14-SF-21	PT Fast derailment at station outside train clearances, towards adjacent line, carriage on its side, secondary collision with PT, no fire
T14-SF-22	PT Fast derailment at station outside train clearances, towards adjacent line, carriage on its side, secondary collision with PT, with fire
T14-SF-23	PT Fast derailment at station outside train clearances, towards adjacent line, carriage on its side, secondary collision with FT carrying non-hazardous goods, no fire
T14-SF-24	PT Fast derailment at station outside train clearances, towards adjacent line, carriage on its side, secondary collision with FT carrying non-hazardous goods, with fire
T10-LCPAU-2	Train strikes normal road vehicle (car/HGV) on AOCL or ABCL crossing, with fire
T10-LCPAU-3	Train strikes large road vehicle above buffer height (low loader, JCB, etc.) on an AOCL or ABCL crossing, no fire
T10-LCPAU-4	Train strikes large road vehicle above buffer height (low loader, JCB, etc.) on an AOCL or ABCL crossing, with fire

Code	Description
T10-LCPAU-5	Train strikes HGV carrying toxic hazardous goods on AOCL or ABCL crossing, no fire
T10-LCPAU-6	Train strikes HGV carrying toxic hazardous goods on AOCL or ABCL crossing, with fire
T10-LCPAU-7	Train strikes HGV carrying flammable hazardous goods on AOCL or ABCL crossing, with fire
T10-LCPRO-2	Train strikes normal road vehicle (car/HGV) on AHB, FP, MB/MCB/CCTV & all UWC (incl. MWL + T & UWC), with fire
T10-LCPRO-3	Train strikes large road vehicle above buffer height (low loader, JCB, etc.) on AHB, FP, MB/MCB/CCTV & all UWC (incl. MWL + T & UWC), no fire
T10-LCPRO-4	Train strikes large road vehicle above buffer height (low loader, JCB, etc.) on AHB, FP, MB/MCB/CCTV & all UWC (incl. MWL + T & UWC), with fire
T10-LCPRO-5	Train strikes HGV carrying toxic hazardous goods on AHB, FP, MB/MCB/CCTV & all UWC (incl. MWL + T & UWC), no fire
T10-LCPRO-6	Train strikes HGV carrying toxic hazardous goods on AHB, FP, MB/MCB/CCTV & all UWC (incl. MWL + T & UWC), with fire
T10-LCPRO-7	Train strikes HGV carrying flammable hazardous goods on AHB, FP, MB/MCB/CCTV & all UWC (incl. MWL + T & UWC) crossing, with fire

Appendix C

Definition and Theorems on Positive Definite and Positive Semi-Definite Matrices

Definition C.1 *Any $n \times n$ symmetric real matrix A is positive definite when $x^T Ax > 0$ for all nonzero vector $x \in R^{n\times 1}$. A symmetric real matrix A is positive semi-definite when $x^T Ax \geq 0$ for all $x \in R^{n\times 1}$; and A is positive definite when $x^T Ax > 0$ for all $x \in R^{n\times 1}$.*

This definition can be found in the books on linear algebra such as [Strang 1986; Herstein and Winter 1988; Zhang 1999]. From this definition, we can develop more properties about the positive definite matrix and the positive semi-definite matrix.

Theorem C.1 *Suppose that A is a $n \times n$ symmetric real matrix of the eigenvalues λ_i, $i = 1, \cdots, n$. The matrix A is positive definite when $\lambda_i > 0$ for $i = 1, \cdots, n$; the matrix A is positive semi-definite $\lambda_i \geq 0$ for $i = 1, \cdots, n$.*

Proofs:

Suppose that $h_i \in R^{n\times 1}$ is the eigenvector associated with λ_i, i.e $Ah_i = \lambda_i h_i$. Base on the linear algebra theory [Strang 1986], h_i $i = 1, \cdots, n$ are orthogonal and normalized, i.e.,

$$h_i^T h_j = \begin{cases} 0, i \neq j \\ 1, i = j \end{cases}$$

Therefore $h_i \quad i=1,\cdots,n$ define a complete base of the $R^{n\times 1}$ space. Therefore any vector $x \in R^{n\times 1}$ can be defined by the linear combination of $h_i \quad i=1,\cdots,n$ as

$$x = \sum_{i=1,n} c_i h_i$$

Base on matrix operation rules [Lay 2003], we have

$$\begin{aligned} x^T Ax &= \sum_{i=1,n} c_i h_i^T \sum_{j=1,n} c_j A h_j \\ &= \sum_{i=1,n} c_i h_i^T \sum_{j=1,n} c_j \lambda_j h_j \\ &= \sum_{i=1,n} \sum_{j=1,n} c_i c_j \lambda_j h_i^T h_j \end{aligned}$$

By the orthogonality and the normalization of h_i, we have

$$x^T Ax = \sum_{i=1,n} c_i^2 \lambda_i$$

By the above formula, $x^T Ax \geq 0$ when $\lambda_i \geq 0$ for $i=1,\cdots,n$. By Definition C.1, the matrix *A* is positive semi-definite and Theorem C.1 is proven.

Theorem C.2 *Suppose that A is a $n\times n$ symmetric real matrix. A is then positive semi-definite when A can be factored as $A = U^T U$, where $U \in R^{n\times n}$.*

Proofs:

By the matrix operation rules [Lay 2003], for any vector $x \in R^{n\times 1}$ we have

$$x^T Ax = x^T \left(U^T U\right) x = \left(Ux\right)^T \left(Ux\right) \geq 0$$

By the linear algebra theory [Strang 1986], we have $Ux = y \in R^{n\times 1}$ and

$$(Ux)^T (Ux) = y^T y = \sum_i y_i^2 \geq 0$$

Therefore the matrix *A* is positive semi-definite and Theorem C.2 is proven.

Theorem C.3 *Suppose that* $A, B \in R^{n \times n}$. *The matrix* $B^T AB$ *is positive semi-definite when A is positive semi-definite.*

Proofs:

By the matrix operation rules [Lay 2003], for any vector $x \in R^{n \times 1}$ we have

$$x^T \left(B^T AB\right)x = \left(x^T B^T\right)A(Bx) = (Bx)^T A(Bx)$$

By the linear algebra theory [Strang 1986], we have $Bx = y \in R^{n \times 1}$. Because *A* is positive semi-definite, we have

$$x^T \left(B^T AB\right)x = y^T Ay \geq 0$$

By Definition C.1, Theorem C.3 is proven.

Theorem C.4 *Suppose that there are m positive semi-definite matrices* $A_i \in R^{n \times n}$, $i = 1, \cdots, m$. *The sum matrix* $A = \sum_{i=1,m} A_i$ *is positive semi-definite; the sum matrix A is positive definite when any* A_i *is positive definite.*

Proofs:

For any vector $x \in R^{n \times 1}$, suppose that $x^T A_i x = \alpha_i$. By Definition C.1 we have $\alpha_i \geq 0$ for $i = 1, \cdots, m$. By the matrix operation rules [Lay 2003], we can have

$$x^T Ax = \sum_{i=1,m} x^T A_i x = \sum_{i=1,m} \alpha_i \geq 0$$

By Definition C.1, the sum matrix A is positive semi-definite. When there exists any positive definite A_i, i.e. $\alpha_i > 0$ for any $x \in R^{n\times 1}$, we have

$$x^T Ax = \sum_{i=1,m} \alpha_i > 0$$

By Definition C.1, the sum matrix A is positive definite. Theorem C.4 is proven.

Appendix D

Reference

Abrahamsson M. (2002). Uncertainty in Quantitative Risk Analysis - Characterisation and Methods of Treatment. Lunds, Lunds University, Sweden.

Adams J. (2001). Risk, Routledge.

Andrews J.D. and Moss T.R. (2002). Reliability and Risk Assessment. London and Bury St Edmunds, UK, Professional Engineering Publishing Limited.

Apolostolakis G.E. (1988). "The Interpretation of Probability in Probabilistic Safety Assessments." Reliabiity Engineering & System Safety **23**: 247-252.

Basak I. (1998). "Probabilistic judgments specified partially in the analytic hierarchy process." European Journal of Operational Research **108**(1): 153-164.

Bazzurro P. and Luco N. (2005). Accounting for uncertainty and correlation in earthquake loss estimation. ICOSSAR, Rome, Italy, Millpress.

Beach L.R. and Swenson R.G. (1966). "Intuitive Estimation of Means." Psychonomic Science **5**: 161-162.

Bedford T. and Cooke R. (2001). Probabilistic Risk Analysis: Foundations and Methods, Cambridge University Press.

Bedford T. and Cooke R.M. (2002). "Vines - A New graphical model for dependent random variables." The Annals of Statistics **30**(4): 1031-1068.

Bishop C.M. (1995). Neural Networks for Pattern Recognition. Oxford, Clarendon Press.

Brown D. and Alexander N. (1991). "The Analysis of the Variance and Covariance of Products." Biometrics **47**: 429-444.

Burden R.L. and Faires D.J. (1997). Numerical Analysis, ITP.

Cagno E., Caron F., et al. (2000). "Using AHP in determining the prior distributions on gas pipeline failures in a robust Bayesian approach." Reliability Engineering & System Safety **67**(3): 275-284.

Cheng D., Bedford T., et al. (2007). Modelling subjective uncertainty in Fault-tree precursors. The 17th Advance in Risk and Reliability Technology Symposium, Loughborough University, UK, Loughborough University.

Clemen R.T., Fischer G.W., et al. (2000). "Assessing Dependence: Some experimental results." Management Science **46**(8): 1100-1115.

Clemen R.T. and Reilly T. (1999). "Correlations and copulas for decision and risk analysis." Management Science **45**(2): 208.

Clemen R.T. and Reilly T. (2001). Making hard decisions : an introduction to decision analysis, Duxbury Press.

Clemen R.T. and Winkler R.L. (1985). "Limits for the precision and value of information from dependent sources." Operations Research **33**: 427-442.

Cooke R.M. (1991). Experts in Uncertainty, Oxford University Press.

Cooke R.M. (1997). "Uncertainty modeling: Examples and issues." Safety Science **26**(1-2): 49-60.

Creswell J.W. (2009). Research Design, SAGE.

De Finetti B. (1974). Theory of Probability: A Critical Introductory Treatment, John Wiley & Sons.

Dennis C. (2005). Profile of safety risk on the UK mainline railway Issue 4, Rail Safety and Standards Board (RSSB), UK.

Dennis C. (2006). Profile of safety risk on the UK mainline railway Issue 5, Railway Safety and Standards Board (RSSB), UK.

Der Kiureghian A. (1987). Multivariate distribution models for structural reliability. Strutural Mechanics in Reactor Technology.

Dillon W.R. and Goldstein M. (1984). Multivariate Analysis: Methods and Applications, John Wiley & Sons.

Duffey M.R. and Van Dorp R.J. (1998). "Risk Aanalysis for Large Engineering Projects: Modeling Cost Uncertainty for Ship Production Activities." Journal of Engineering Valuation and Cost Analysis **2**: 285-301.

Evans A.W. (2003). Transport fatal accidents and FN-curves: 1967-2001. London, Her Majesty's Stationery Office.

Evans M., Nicholas H., et al. (2000). Statistical Distributions, Wiley.

Ferson S. and Hajagos J.G. (2005). "Varying correlation coefficients can underestimate uncertainty in probabilistic models." Reliability Engineering & System Safety **In Press, Corrected Proof**.

Fox C. (1961). "The G and H Functions as Symmetrical Fourier Kernels." Transactions of the American Mathematical Society **98**: 395-429.

Frees E.W. and Valdez E.A. (1998). "Understanding relationships using copulas." North American Acturial Journal **2**(1): 1-25.

Frey C.H. and Patil S.R. (2001). Identification and Review of Sensitivity Analysis Methods, Civil Engineering Department, North Carolina State University.

Garthwaite P.U., Kadane J.B., et al. (2005). "Statistical Methods for Eliciting Probability Distributions." Journal of the American Statistical Association **100**(470): 680-700.

Gelman A., Carlin J.B., et al. (2003). Bayesian Data Analysis, Chapman & Hall/CRC.

Ghosh S. and Henderson S. (2002). PROPERTIES OF THE NORTA METHOD IN HIGHER DIMENSIONS. Proceedings of the 2002 Winter Simulation Conference.

Ghosh S. and Henderson S. (2003). "Behavior of the NORTA method for correlated random vector generation as the dimension increases." ACM Transactions on Modeling and Computer Simulation (TOMACS) **13**(3): 276-294.

Ghosh S. and Henderson S.G. (2002). "Chessboard distributions and random vectors with specified marginals and covariance matrix." Operations Research **50**(5): 820-834.

Glen A.G., Leemis L.M., et al. (2004). "Computing the distribution of the product of two continuous random random variables." computational statistics & Data analysis **44**: 451-464.

Gokhale D.V. and Press S.J. (1982). "Assessment of a Prior Distribution for the Correlation coefficiet in a Bivariate Normal distribution." Journal of the Royal Statistical Society A: Statistics in Society **145**(2): 237-249.

Harrison C. (2004). The Use of Empirical Bayes Methods in the Rail Safety & Standards Board (RSSB) Safety Risk Model. The 27th ESReDA Seminar: Assembling evidence of reliability: The reliability case as a decision support tool, Glasgow, Scotland, IPSC.

Harrison C., Griffin D., et al. (2008). Modelling Uncertainty in the RSSB Safety Risk Model. The 35th ESReDA Seminar: Uncertainty in Industrial Practice - Generic best practices in uncertainty treatment, Marseille, France.

Haugom G.P., Rikheim H., et al. (1990). Hydrogen applications risk acceptance criteria and risk assessment methodology, DNV, Norway.

Health and Safety Executive (1992). The tolerability of risk from nuclear power stations, Her Majesty's Stationery Office.

Helton J.C. (1999). "Uncertainty and sensitivity analysis in performance assessment for the Waste Isolation Pilot Plant." Computer Physics Communications **117**(1-2): 156-180.

Helton J.C., Anderson D.R., et al. (1996). "Uncertainty and sensitivity analysis results obtained in the 1992 performance assessment for the waste isolation pilot plant." Reliability Engineering & System Safety **51**(1): 53-100.

Helton J.C., Anderson D.R., et al. (2000). "Conceptual structure of the 1996 performance assessment for the Waste Isolation Pilot Plant." Reliabiity Engineering & System Safety **69**: 151-165.

Helton J.C., Bean J.E., et al. (1996). "Uncertainty and sensitivity analysis for gas and brine migration at the Waste Isolation Pilot Plant: Permeable shaft with panel seals." Journal of Hazardous Materials **45**(2-3): 107-139.

Helton J.C., Bean J.E., et al. (1997). "Uncertainty and sensitivity analysis for gas and brine migration at the Waste Isolation Pilot Plant: permeable shaft without panel seals." Reliability Engineering & System Safety **57**(3): 299-316.

Helton J.C. and Davis F.J. (2000). Sampling-based methods. Sensitivity Analysis. A. Saltelli, K. Chan and E. M. Scott.

Helton J.C. and Davis F.J. (2002). "Illustration of sampling-based methods for uncertainty and sensitivity analysis." Risk Analysis **22**(3): 591-622.

Helton J.C. and Davis F.J. (2003). "Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems." Reliability Engineering & System Safety **81**(1): 23-69.

Helton J.C., Davis F.J., et al. (2005). "A comparison of uncertainty and sensitivity analysis results obtained with random and Latin hypercube sampling." Reliability Engineering & System Safety **89**(3): 305-330.

Helton J.C., Johnson D., et al. (2004). "An exploration of alternative approaches to the representation of uncertainty in model predictions." Reliabiity Engineering & System Safety **85**: 39-71.

Helton J.C., Johnson J.D., et al. (1995). "Uncertainty and sensitivity analysis of chronic exposure results with the MACCS reactor accident consequence model." Reliability Engineering & System Safety **50**(2): 137-177.

Helton J.C., Johnson J.D., et al. (1995). "Uncertainty and sensitivity analysis of food pathway results with the MACCS reactor accident consequence model." Reliability Engineering & System Safety **49**(2): 109-144.

Helton J.C., Johnson J.D., et al. (2006). "Survey of sampling-based methods for uncertainty and sensitivity analysis." Reliability Engineering & System Safety **91**(10-11): 1175-1209

Helton J.C., Johnson J.D., et al. (1995). "Uncertainty and sensitivity analysis of early exposure results with the MACCS reactor accident consequence model." Reliability Engineering & System Safety **48**(2): 91-127.

Helton J.C., Martell M.A., et al. (2000). "Characterization of subjective uncertainty in the 1996 performance assessment for the Waste Isolation Pilot Plant." Reliability Engineering & System Safety **69**(1-3): 191-204.

Herstein I.N. and Winter D.J. (1988). Matrix theory and linear algebra, Macmillan.

HSE(Health & Safety Executive) (2001). Reducing risks, protecting people. London, Her Majesty's Stationery Office.

Hughes W.R. (1993). "Consistent utility and probability assessment using AHP methodology." Mathematical and Computer Modelling **17**(4-5): 171-177.

Iman R., Johnson M.E., et al. (2002). "Assessing hurricane effects. Part 2. Uncertainty analysis." Reliability Engineering and System Safety **78**(2002): 147-155.

Iman R.L. and Conover W.J. (1982). "A distribution-Free Approach to Inducing Rank Correlation among input variables." Communications in Statistics Part B: Simulation and Computation **11**(3): 311-334.

Iman R.L. and Helton J.C. (1991). "The repeatability of Uncertainty and Sensitivity analyses for complex Probabilistic risk assessments." Risk Analysis **11**(4): 591-606.

Isograph (2004). FaultTree+ V11.0 Dynamic Link Library V1 User Manual.

Isograph (2005). FaultTree+ for Windows V11.0.

Jelen B. and Syrstad T. (2004). VBA and Macros for Microsoft Excel, QUE.

Johnson N.L. and Balakrishnan S.K. (1994). Continuous univariate distributions, Wiley & Sons.

Jolliffe I.T. (2002). Principal Component Analysis, Springer.

Jonkman S.N., Van Gelder P.H.A.J.M., et al. (2003). "An overview of quantitative risk measures for loss of life and economic damage." Journal of Hazardous Materials **A99**: 1-30.

Jouini M.N. and Clemen R.T. (1996). "Copula models for aggregating expert opinions." Operations Research **44**: 444-457.

Kadane B.J. (1980). Predictive and Structural Methods for Eliciting Prior Distributions. Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jefferys. A. Zellner. Amsterdam, Holland.

Kadane J. and Wolfson L.J. (1998). "Experiences in elicitation." Statistician **47**(1): 1-20.

Kaplan S. and Garrick B.J. (1981). "On the Quantitative Definition of Risk." Risk Analysis **1**(1): 11-27.

Keefer L.D. and Bodily E.S. (1983). "Three-point approximations for continuous random variables." Management Science **29**(5): 595-609.

Kraan B.C.P. and Cooke R.M. (2000). "Uncertainty in compartmental models for hazardous materials -- a case study." Journal of Hazardous Materials **71**(1-3): 253-268.

Kruskal W.H. (1958). "Ordinal measures of association." Journal of the American Statistical Association **53**(284): 814-861.

Krzanowski W.J. (1988). Principles of Multivariate Analysis, Oxford University Press.

Kumamoto H. and Henley E.J. (1996). Probabilistic Risk Assessment and Management for Engineers and Scientists, IEEE PRESS.

Kurowicka D. and Cooke R.M. (2006). Uncertainty Analysis with High Dimensional Dependence Modelling, Wiley.

L'Ecuyer P. (1998). Random number generation. Handbook of Simulation. J. Banks, John Wiley & Sons.

Lauridsen K., Kozine I., et al. (2002). Assessment of Uncertainties in Risk Analysis of Chemical Establishments, Riso National Laboratory, Roskilde, Denmark**:** 52.

Law A.M. and McComas M.G. (1999). EXPERTFIT: TOTAL SUPPORT FOR SIMULATION INPUT MODELING. Winter Simulation Conference.

Lay D.C. (2003). Linear Algebra and its applications, Pearson Education.

Li Y. and Ellingwood B.R. (2006). "Hurricane damage to residential construction in the US: Importance of uncertainty modeling in risk assessment." Engineering Structures **28**(6): 1009-1018.

Lichtenstein S. and Fischhoff B. (1980). "Training for calibration." Organizational Behavior and Human Performance **26**(2): 149-171.

Lindley D. (2000). "The Philosophy of Statistics." Journal of the Royal Statistical Society D: Statistician **49**(3): 293-337.

Lomnicki Z.A. (1967). "On the Distribution of Products of Random Variables." Journal of the Royal Statistical Society. Series B **29**(3): 513-524.

Marsh D.W.R. and Bearfield G. (2008). "Generalizing event trees using Bayesian networks." Journal of Risk and Reliability **222**: 105-114.

McKay M.D. and Beckman R.J. (1979). "A comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code." Technometrics **21**(2): 239-245.

Mclachlan G. and Peel D. (2000). Finite Mixture Models, John Wiley & Sons.

Merkhofer M.W. (1987). "Quantifying Judgemental Uncertainty: Methodology, Experience, and Insights." IEEE Transactions on system, Man, and Cybernetics **SMC-17**(5): 741-752.

Modarres M. (2006). Risk Analysis in Engineering: techniques, tools, and trends, Taylor & Francis.

Monti S. and Carenini G. (2000). "Dealing with the expert inconsistency in probability elicitation." IEEE Transactions on Knowledge and Data Engineering **12**(4): 499-508.

Morgan G.M. and Henrion M. (1990). Uncertainty - A guide to dealing with Uncertainty in Quantative Risk and Policy Analysis, Cambridge University Press.

Mula J., Poler R., et al. (2006). "Models for production planning under uncertainty: A review." International Journal of Production Economics **103**(1): 271-285.

Murphy A.H. and Winkler R.L. (1974). "Credible interval temperature forecasting: some experimental results." Monthly Weather Review **102**: 784-794.

Nadarajah S. and Ali M.M. (2006). "The distribution of sums, products and ratios for Lawrance and Lewis's bivariate exponential random variables." Computational Statistics & Data Analysis **50**: 3449-3463.

Nelsen R.B. (2006). An introduction to copulas. New York, Springer.

Nilsen T. and Aven T. (2003). "Models and model uncertainty in the context of risk analysis." Reliabiity Engineering & System Safety **79**: 309-317.

Nilsen T., Gudmestad O.T., et al. (1998). "Utilisation of principles from structural reliability in quantitative risk analysis: example from an offshore transport problem." Reliability Engineering & System Safety **61**(1-2): 127-137.

O'Hagan A. (1998). "Eliciting Expert Beliefs in Substantial Practical Applications." The Statistician **47**(1): 21-35.

O'Hagan A., Buck C.E., et al. (2006). Uncertainty Judgement: Eliciting Experts' Probabilities, John Wiley & Sons.

ORACLE (2008, 2008). "CRYSTAL BALL TUTORIAL." from http://www.decisioneering.com/tutorial.html.

Papazoglou I.A. (1998). "Mathematical foundations of event trees." Reliabiity Engineering & System Safety **61**: 169-183.

Park K.S. and Lee J.i. (2008). "A new method for estimating human error probabilities: AHP-SLIM." Reliabiity Engineering & System Safety **93**: 578-587.

Parry G.W. (1996). "The characterization of uncertainty in probabilitic risk assessment of complex systems." Reliabiity Engineering & System Safety **54**: 119-126.

Pearson E.S. and Tukey J.W. (1965). "Approximate means and standard deviations based on distances between percentage points of frequency curves." Biometrika **52**(3, 4): 533-546.

Peterson C. and Miller A. (1964). "Mode, median, and mean as optimal strategies." Journal of Experimental Psychology **68**(4): 363-367.

Peterson C.R., Snapper K.J., et al. (1972). "Credible interval temperature forecasts." Bulletin of the American Meteorological Society.

Pidd M. (2003). Tools for Thinking: Modelling in Management Science, John Wiley & Sons.

Ponterotto J.G. (2005). "Qualitative research in counseling psychology: A primer on research paradigms and philosophy of science." Journal of Counseling Psychology **52**(2): 126-136.

Rabitz H. (1989). "System analysis at molecular scale." Science **246**(221-226).

Ravinder H.V., Klenmuntz D.N., et al. (1988). "The reliability of subjective probabilitties obtained through decomposition." Management Science **34**: 186-199.

Robinson D.G. (1998). A survey of probabilistic methods used in reliability, risk and uncertainty analysis, Aandia National Laboratories.

Ross S.M. (2003). Introduction to Probability Models, Academic Press.

RSSB (2004). Road Vehicle Level Crossings Special Topic Report, Rail Safety and Standard Board.

Rushdi A.M. (1985). "Uncertainty Analysis of Fault-Tree Outputs." IEEE Transactions on Reliability **34**(5): 458-462.

Saaty T.L. (1977). "A scaling method for priorities in hierarchical structures." Journal of Mathematical Psychology **15**(3): 234-281.

Saaty T.L. (1980). The analytic hierarchy process : planning, priority setting, resource allocation. New York, McGraw-Hill International Book Co.

Sallaberry C.J., Helton J.C., et al. (2006). Extension of Latin Hypercube Samples with Correlated Variables, Sandia National Laboratories.

Salo J., El-Sallabi H., et al. (2006). "The Distribution of the Product of Independent Rayleigh Random Variables." IEEE Transactions on Antennans and propagation **54**(2): 639-643.

Saltelli A., Chan K., et al. (2000). Sensitivity Analysis, John Wiley & Sons.

Saltelli A.T.S.C.F.R.M. (2004). Sensitivity Analysis in Practice, John Wiley & Sons.

Schaefer R.E. and Borcherding K. (1973). "The assessment of subjective probability distributions: A training experiment." Acta Psychologica **37**(2): 117-129.

Scheuer E.M. and Stoller D.S. (1962). "On the Generation of Normal Random Vectors." Technometrics **4**: 278-281.

Seaver D.A., Von Winterfeldt D., et al. (1978). "Eliciting subjective probability distributions on continuous varaibles." Organizational Behavior and Human Decision Processes **21**: 379-391.

Shephard G.G. and Kirkwood C.W. (1994). "Managing the judgmental probability elicitation process: a casestudy of analyst/manager interaction." IEEE Transactions on Engineering Management **41**(4): 414-425.

Smith A.E., Ryan J.S., et al. (1992). "The effect of neglecting correlations when propogating uncertainty and estimating the population distribution of risk." Risk Analysis **12**: 467-474.

Sobol I.M. (1974). The Monte Carlo Method, The University of Chicago Press.

Soll J.B. and Klayman J. (2004). "Overconfidence in Interval Estimates." Journal of Experimental Psychology: Learning, Memory, and Cognition **30**(2): 299-314.

Spencer J. (1961). "Estimating averages." Ergonomics **4**: 317-328.

Spencer J. (1963). "A further study of estimaitng averages." Ergonomics **6**(3): 255-265.

Springer M.D. (1979). The Algebra of Random Variables. New York, Wiley.

Springer M.D. and Thompson W.E. (1966). "The Distribution of Products of Independent Random Variables." SIAM Journal on Applied Mathematics **14**(3): 511-526.

Springer M.D. and Thompson W.E. (1970). "The distribution of product of Beta, Gamma and Gaussian random variables." SIAM Journal on Applied Mathematics **18**(4): 721-737.

Stein M. (1987). "Large Sample Properties of Simulations Using Latin Hypercube Sampling." Technometirc **29**(2): 143-151.

Strang G. (1986). Linear Algebra and its Applications. San Diego, Harcourt Brace Jovanovich.

Szwed P., Van Dorp R.J., et al. (2006). "A Bayesian Paired Comparision Approach for Relative Accident Probability Assessment with Covariate Information." European Journal of Operational Research **169**(1): 157-177.

Titterington D.M.S.A.F.M.M.U.E. (1985). Statistical Analysis of Finite Mixture Distributions, John Wiley & Sons.

Tong Y.L. (1990). The Multivariate Normal Distribution. New York, Springer-Verlag.

Van Dorp R.J. (2005). "Statistical dependence through common risk factors: With applications in uncertainty analysis." European Journal of Operational Research **161**(1): 240-255.

Van Kampen N.G. (1992). Stochastic processes in physics and chemistry. New York, North-Holland.

Versteeg M.F. (1988). "External safety policy in the Netherlands: an approach to risk management." Journal of Hazardous Materials **17**(2): 215-221.

Walls L. and Quigley J. (2001). "Building prior distributions to support Bayesian reliability growth modelling using expert judgement." Reliability Engineering & System Safety **74**(2): 117-128.

Watkins D.S. (1991). Fundamentals of Matrix Computations, John Willey & Sons.

Weisstein E.W. "Normal Product Distribution." from http://mathworld.wolfram.com/NormalProductDistribution.html

Wilkinson J.H. (1965). The Algebraic Eigenvalue Problem, Oxford University Press.

Winkler R.L. (1980). Prior Information, Predictive Distributions, and Bayesian Model Building. Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jeffreys. A. Zellner. Amsterdam, Holland.

Winkler R.L. (1996). "Uncertainty in probabilistic risk assessment." Reliability Engineering & System Safety **54**(2-3): 127-132.

Winman A., Hansson P., et al. (2004). "Subjective Probability Intervals: How to Reduce Overconfidence by Interval Evaluation." Journal of Experimental Psychology: Learning, Memory, and Cognition **30**(6): 1167-1175.

Yi W. and Bier V.M. (1998). "An application of Copulas to Accident Precursor Analysis." Management Science **44**(12): S257-S270.

Zhang F. (1999). Matrix theory, Springer.