RSP-0296

Final Report

Statistical Analysis of Common Cause Failure Data to Support Safety and Reliability Analysis of Nuclear Plant Systems

Professor Mahesh Pandey
University of Waterloo
Waterloo, ON N2L 3G1

Disclaimer

Abstract

This report describes the findings of a project "Statistical Analysis of Common Cause Failure Data to Support Safety and Reliability Analysis of Nuclear Plant Systems for the CNSC" under the contract No. 87055-12-0221.

Analysis of Common Cause Failures (CCF) is an important element of the Probabilistic Safety Assessment (PSA) of systems important to safety in a nuclear power plant. Based on the conceptualization of the CCF event, many probabilistic models have been developed in the literature. This Report provides a comprehensive review of CCF modeling techniques, which shows that the a modern method, called "General Multiple Failure Rate Model", is the most suitable method for probabilistic modeling of CCF events. Therefore, the GFMR is described in detail in the report and adopted for the case studies. To estimate the parameters of the GMFR model, the Empirical Bayes (EB) method is adopted. The report describes the data mapping methods and the EB method for combining data from different component groups and plants in the statistical estimation.

This project presents detailed case studies to illustrate the data mapping and EB method. The case studies are based on CCF data for motor operated valves (MOVs). These case studies serve as templates to analyze CCF data from other safety systems. The report provides analysis methods to the CNSC staff to analyze CCF rates and evaluate the adequacy of input data used in the PSA of Canadian plants.

This project demonstrates the development of the capacity to analyze CCF data in line with best international practices.

Table of Contents

Disclaimer ii
Abstract iii
Table of Contents iv
List of Figures vii
List of Tables viii
Notations x
Section 1: Introduction 11
1.1 Background 11
1.2 Objective and Scope 12
1.3 Organization 12
Section 2: Basic Concepts of CCF Modeling 13
2.1 Common Cause Failures (CCF) 13
2.2 Homogeneous Poisson Process 13
2.3 General Multiple Failure Rate Model (GMFR) 14
2.4 Alpha Factors 15
2.5 Illustrative Example 15
2.5.1 1oo3 System 16
2.5.2 2oo3 System 17
2.5.3 3oo3 System 18
2.5.4 Example 19
2.6 Summary 19
Section 3: Data Mapping Methods 21
3.1 Introduction 21
3.2 Mapping Down of Data 21
3.3 Mapping Up of Data 24
3.3.1 NUREG-4780 Procedure 24
3.3.2 Method of Vaurio (2007) 26
3.4 Summary 27

Section 4: Statistical Estimation of Failure Rates 28
4.1 Introduction 28
4.2 Maximum Likelihood Estimation (MLE) 28
4.3 Bayesian Method 29
4.3.1 Basic Concepts 29
4.3.2 Empirical Bayes Method 30
4.4 Numerical Example 31
4.5 Summary 36
Section 5: Case Studies 37
5.1 Introduction 37
5.2 Example 1: Motor Operated Valves 37
5.2.1 MLE without Data Mapping 38
5.2.2 MLE with Data Mapping 38
5.2.3 Empirical Bayes (EB) without Data Mapping 39
5.2.4 EB with Data Mapping (Multiple Priors) 41
5.2.5 EB with Data Mapping (Single Prior) 42
5.2.6 Comparison of Results: Case Study 1 44
5.3 Example 2: Larger MOV Data 48
5.3.1 MLE without Data Mapping 48
5.3.2 MLE with Data Mapping 49
5.3.3 EB without Data Mapping 50
5.3.4 EB with Data Mapping (Multiple Priors) 51
5.3.5 EB with Data Mapping (Single Prior) 52
5.3.6 Comparison of Results 53
5.4 Summary 57
Section 6: Summary and Conclusions 58
6.1 Summary 58
6.2 Conclusions 58
6.3 Recommendations 59

Appendix A Review of CCF models 60
Appendix B Derivation of Vaurio's EB estimator 69
References 72

List of Figures

Figure 2.1: A parallel system with 3 components 15

Figure 2.2: Cut set of 1oo3 system 16

Figure 2.3: Cut set of 1oo3 system with CCF 17

Figure 2.4: Cut sets of 2oo3 system 17

Figure 2.5: : Cut sets of 2oo3 system with CCF 18

Figure 2.6: Cut sets of 3oo3 system with CCF 18

Figure 2.7: Mapping down CCF rates from CCG 2 to 1 22

Figure 2.8: Mapping up from CCG 1 to 2. 24

Figure 4.1: Comparison of mean and SD of failure rates (MLE and EB methods) 33

Figure 4.2: Distribution of the failure rate: component 1 34

Figure 4.3: Distribution of the failure rate: component 4 34

Figure 4.4: Distribution of the failure rate: component 8 35

Figure 5.1: Failure rate distribution without data mapping: case study 1 40

Figure 5.2: Failure rate distributions with data mapping (multiple priors) 42

Figure 5.3: Failure rate distributions with data mapping (single prior) 43

Figure 5.4: Comparison of mean failure rates: case study 1 45

Figure 5.5: Comparison of standard deviations of failure rates by different methods 46

Figure 5.6: Comparison of alpha factors by different methods 47

Figure 5.7: Failure rate distributions without data mapping: case study 2 50

Figure 5.8: Failure rate distributions with data mapping (multiple priors) 52

Figure 5.9: Failure rate distributions with data mapping (single prior) 53

Figure 5.10: Comparison of estimates of mean failure rates (case study 2) 54

Figure 5.11: Comparison of estimates of SD of failure rates: case study 2 55

Figure 5.12: Comparison of alpha factors: case study 2 56

List of Tables

Table 2.1: Failure rates of a component 19
Table 2.2: Reliability of a k-out-of-3 system 19
Table 2.3: Mapping down formulas for CCG up to 4 (NUREG-4780). 23
Table 2.4: Formulas for mapping up failure rates for CCG 1 to 4 (NUREG-4780). 25
Table 4.1: EB method for estimating the gamma prior from data 30
Table 4.2: MLE and EB estimates of failure rates 31
Table 4.3: The median and 95th percentile of the failure rates from posterior distribution 35
Table 5.1: CCF Data: case study 1 37
Table 5.2: MLE results: case study 1 38
Table 5.3: CCF data mapped to CCG of 4 39
Table 5.4: MLE results with data mapping: case study 1 39
Table 5.5: Posterior Mean and SD of failure rate without data mapping 40
Table 5.6: Number and exposure times of 1oo4 failures after data mapping 41
Table 5.7: Posterior Mean and SD of failure rates with data mapping 41
Table 5.8: Parameters of Priors (mapped data) 41
Table 5.9: Posterior mean and of failure rates with data mapping 43
Table 5.10: Mean failure rates (per month) by different methods: case study 1 44
Table 5.11: SD of failure rates by different methods: case study 1 45
Table 5.12: Alpha factors by different method: case study 1 46
Table 5.13: CCF data: case study 2 48
Table 5.14: MLE results without data mapping: case study 2 48
Table 5.15: CCF data mapped to CCG of 4: case study 2 49
Table 5.16: MLE results after data mapping: case study 2 49
Table 5.17: Posterior Mean and SD of failure rate without data mapping 50
Table 5.18: Number and exposure times of 1oo4 failures after data mapping 51
Table 5.19: Posterior mean and SD of failure rate with data mapping 51
Table 5.20: Parameters of priors (mapped data): case study 2 52
Table 5.21: Posterior mean and SD of failure rates (mapped data, single prior) 53

Table 5.22: Mean failure rates (/month) by different methods (CCG=4)............................... 54
Table 5.23:SD of failure rates (/month) by different methods: case study 2)......................... 55
Table 5.24: Alpha factors by different methods: case study 2.. 56
Table 5.25: Failure data of diesel generators in SKI Report 2009:07. 64
Table 5.26: Failure rates of a k-out-of-4 system calculated by different models. 64

Notations

AF	Alpha Factor
BF	Beta Factor
BFR	Binomial Failure Rate
CCG	Common Cause Component Group
CCF	Common Cause Failure
CV	Coefficient of Variation
EB	Empirical Bayes
EDG	Emergency Diesel Generator
FO	Failure to Open
HPP	Homogeneous Poisson Process
*k*oo*n*	*k* out of *n* components must be reliable to ensure system reliability
MGL	Multiple Greek Letter
MOV	Motor-operated Valve
MSE	Mean Squared Error
NUREG	Nuclear Regulatory Commission
PDF	Probability Density Function
SD	Standard Deviation
USNRC	United States Nuclear Regulatory Commission

Section 1:
Introduction

1.1 Background

In the context of Probabilistic Safety Assessments (PSA), Common Cause Failure (CCF) events arc a subset of dependent events in which two or more components fail within a short interval of time as a result of a shared (or common) cause. Common cause events are highly relevant to PSA due to their potential adverse impact on the safety and availability of critical safety systems in the nuclear plant. An accurate estimation of CCF rates is therefore important for a realistic PSA of plant safety systems.

Due to lack of data, CCF rates/probabilities were estimated by expert judgment in early years of PSA. Over the years as the data were collected by the utilities and regulators world-wide, more formal statistical analysis methods for data analysis emerged to derive improved estimates of CCF rates. The estimates of CCF rates in line with the operating experience should be used in PSA in place of generic or expert judgment estimates.

International Common Cause Failure Data Exchange (ICDE) is a concerted effort by undertaken by many countries to compile the CCF event data in a consistent manner. CCF event data are compiled by a third party (OECD) in a database with a copy maintained by the CNSC. The project aims to utilize the data compiled by ICDE for the estimation of CCF rates.

This project first aims to understand the current status of CCF modeling techniques and international best practices. The underlying probability theory is carefully evaluated to facilitate the training of industry professionals in this area. For example, concepts underlying a current estimation method, Empirical Bayes (EB) method, are clearly illustrated through examples.

This project presents detailed case studies to illustrate the data mapping and EB method for data analysis. The case studies are based on CCF data for motor operated valves (MOVs). These case studies serve as templates to analyze CCF data from other safety systems. The report provides analysis methods to the CNSC staff to analyze CCF rates and evaluate the adequacy of input data used in the PSA of Canadian plants.

1.2 Objective and Scope

The main objective of this project is to analyze probabilistic modeling of CCF events and present statistical methods for the parameter estimation. Since CCF rates are used as input to the Probabilistic Safety Assessment (PSA), the application of logically sound model discussed in the report would improve the confidence in the final results of PSA.

The scope of the project is as follows:

- Perform a comparative study between the different methodologies used world-wide to assess the CCF parameters. These methods include: (1) Alpha factor, (2) Beta factor, (3) Multiple Greek letter, and (4) Binomial failure rate.
- Explore the development of other modern methods of statistical analysis for the estimation of CCF rates.
- Analyze statistically CCF events compiled in the ICDE database maintained by the CNSC.
- Develop the calculation models of the CCF parameters for different population size and different testing schemes (staggered and non-staggered testing)
- Present case studies to illustrate the analysis of CCF data. The CCF events will include the following components: Batteries, Heat Exchangers, Diesel generators, Motor operated valves (MOV), Safety relief Valves (SRV), and Check valves.
- Investigate the technical aspects of the data mapping and Empirical Bayes (EB) method for statistical estimation of CCF rates and Alpha factors.

1.3 Organization

This draft report is divided into 6 Sections and 2 Appendices. In Section 2, probabilistic concepts underlying the CCF modeling techniques are discussed. Section 3 reviews the data mapping methods and Section 4 describes the Empirical Bayes (EB) method used in ICDE project for combining data from various plants. Section 5 presents two case studies based on the CCF data for motor operated valves (MOVs). In the last Section, conclusions and recommendations are presented. The first Appendix presents the literature review regarding method of CCF modeling. The second Appendix explains the EB method used by the Nordic members of the ICDE project.

Section 2:
Basic Concepts of CCF Modeling

2.1 Common Cause Failures (CCF)

Nuclear power plant is a complex technological system which requires high level of operational safety and reliability. Safety systems include motor-operated valves (MOV), emergency diesel generators (EDG), water pumps, power batteries and many other devices depending on the functional requirement. In order to achieve high reliability, redundancy of various orders is added to safety systems.

Most of the common causes can be classified into four types, namely, hardware equipment failure, human error during operation, environmental stress applied to components, and external events that causes stress (Mosleh *et al*., 1989).

The key objective of this Section is to introduce the basic concepts and terminology associated with the modeling of common cause failures. The "General Multiple Failure Rate" (GMFR) model is described which is the basis of the CCF analysis in ICDE project.

2.2 Homogeneous Poisson Process

The probabilistic basis for CCF modeling is that the occurrences of failures in a single component are modeled as the homogeneous Poisson process (HPP). It means: (1) failures are purely random without any trend due to ageing, (2) occurrences of failure events are independent of each other, and (3) after a failure component is renewed to its original "as new" condition.

In a time interval (0, *t*), the number of failures are given by the Poisson distribution as

$$P[N(t) = k] = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \tag{2.1}$$

Note that the parameter λ denotes the failure rate, defined as the average number of failures per unit time. The reliability, i.e., no failure in a time interval (0, *t*) is given s

$$R(t) = P[N(t) = 0] = e^{-\lambda t} \tag{2.2}$$

Thus, the failure rate is a single parameter that determines the component reliability. Based on component reliabilities, the system reliability can be evaluated, as illustrated by the following example.

2.3 General Multiple Failure Rate Model (GMFR)

Historically, several conceptual probabilistic models of CCF events have been presented in the literature and they are reviewed in Appendix A. The General Multiple Failure Rate (GMFR) model has been a new and widely accepted model of CCF events. This model is also used in ICDE project.

The basic idea is that a failure observed at the system level involving either a single component failure or a failure of *k*-components is caused by an external shock generated by independent HPPs. In short, n independent HPPs are generating external shocks that cause CFF events of various multiplicities. These HPPs are mutually exclusive, i.e., one HPP is in action at any given time.

In an n-component system, failure event data are described using the following parameters ($k = 1,2, \ldots, n$):

$N_{k/n}$ = Number of failures involving any k components

T = Operation time of the system in which $N_{k/n}$ failures occurred

$\Lambda_{k/n}$ = Failure rate of the HPP causing k out of n component failures

Λ_n = Sum of failure rates of all n HPP = $\sum_{k=1}^{n} \quad _{k/n}$

The failure events that are observed at the system level are caused by component failures. So it is assumed that component failures are caused by shocks modeled as HPPs. Since component failures produce system failure events, the failure rates at the system and component levels are related.

Define

$\lambda_{k/n}$ = Failure rate of an HPP causing a CCF involving k specific components

The failure rates at the system and component levels are then related as

$$\Lambda_{k/n} = \binom{n}{k} \lambda_{k/n} = \frac{N_{k/n}}{T} \tag{2.3}$$

$$\Lambda_n = \sum_{k=1}^{n} \binom{n}{k} \lambda_{k/n} = \sum_{k=1}^{n} \Lambda_{k/n} \tag{2.4}$$

2.4 Alpha Factors

Based on the GMFR model, the alpha factors are defined as the following ratios of system CCF rates

$$\alpha_{k/n} = \frac{\Lambda_{k/n}}{\Lambda_n} \tag{2.5}$$

2.5 Illustrative Example

Consider a redundant system with three components, as shown in Figure 2.1. The failure of an i^{th} component is modeled as an HPP with the failure rate λ_i .

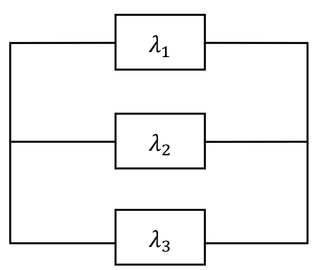


Figure 2.1: A parallel system with 3 components

The probability of failure $P_i(T)$ and reliability $R_i(T)$ of the component i in the time interval $(0, T)$ are

$$P_i(T) = 1 - e^{-\lambda_i T} \tag{2.6}$$

$$R_i(T) = e^{-\lambda_i T} \tag{2.7}$$

2.5.1 1oo3 System

Let's define a *one-out-of-three* system, which means at least one component should be reliable in order to assure that the system is reliable. The cut set of the system failure shown below.

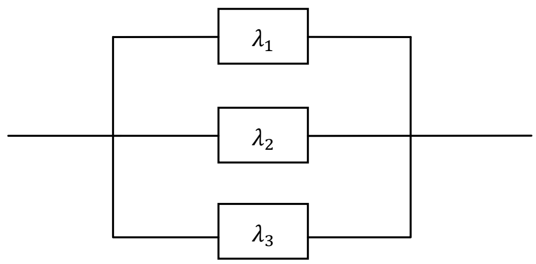


Figure 2.2: Cut set of 1oo3 system

First consider that the system involves only independent failures of the components. In order to fail the system, all the three components must fail. Thus, the probability of failure of the system is

$$P_s(T) = P_1 P_2 P_3 = \left(1 - e^{-\lambda_1 T}\right)\left(1 - e^{-\lambda_2 T}\right)\left(1 - e^{-\lambda_3 T}\right) \tag{2.8}$$

Since the symmetry assumption is often adopted in the failure analysis, all the components would be considered as similar/identical. Therefore, the component failure rates are the same, denoted as $\lambda_{1/3} = \lambda_1 = \lambda_2 = \lambda_3$. Equation (2.8) can be simplified as

$$P_s(T) = \left(1 - e^{-\lambda_{1/3} T}\right)^3 \tag{2.9}$$

Therefore, the reliability of the above system is

$$\begin{aligned} R_s(T) &= 1 - \left(1 - e^{-\lambda_{1/3} T}\right)^3 \\ &= 3e^{-\lambda_{1/3} T} - 3e^{-2\lambda_{1/3} T} + e^{-3\lambda_{1/3} T} \end{aligned} \tag{2.10}$$

If a CCF of the system is postulated, it means that all the three components should fail simultaneously with the failure rate λ_{123}. The cut set of the system failure with CCF is shown below.

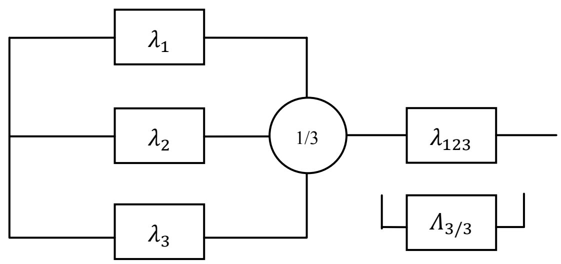


Figure 2.3: Cut set of 1oo3 system with CCF

Since the CCF event is in series with the original cut set, each part should be reliable for the sake of system reliability. In this case, system reliability is modified as

$$R_s'(T) = R_s(T)\left(e^{-\lambda_{123}T}\right)$$
$$= \left(3e^{-\lambda_{1/3}T} - 3e^{-2\lambda_{1/3}T} + e^{-3\lambda_{1/3}T}\right)\left(e^{-\lambda_{3/3}T}\right) \quad (2.11)$$

2.5.2 2oo3 System

2oo3 system means that at least two components must be reliable to ensure the system reliability. The cut sets for the system failure are shown below.

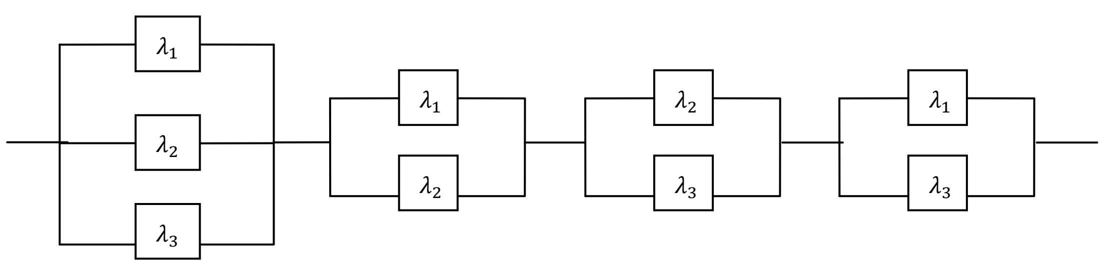


Figure 2.4: Cut sets of 2oo3 system

Using the symmetry assumption, the system reliability without CCF is given as

$$R_s(T) = 3R_1^2P_1 + R_1^3 = 3e^{-2\lambda_{1/3}T}\left(1 - e^{-\lambda_{1/3}T}\right) + e^{-3\lambda_{1/3}T} \quad (2.12)$$

$$= 3e^{-2\lambda_{1/3}T} - 2e^{-3\lambda_{1/3}T}$$

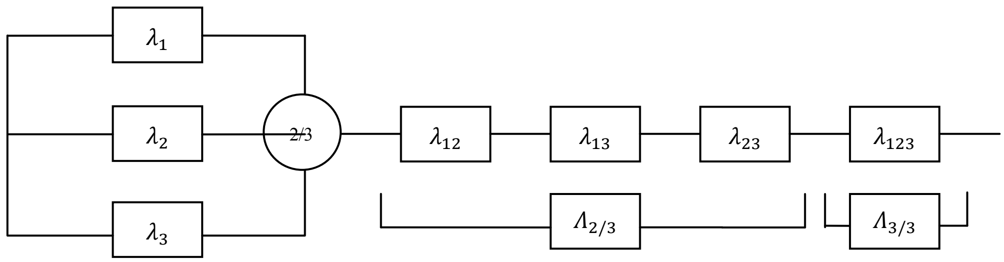


Figure 2.5: : Cut sets of 2oo3 system with CCF

Taking into account of CCF events, the cut sets are modified as shown above. The reliability is given as

$$\begin{aligned} R_s'(T) &= R_s(T)\left(e^{-\lambda_{12}T}\right)\left(e^{-\lambda_{13}T}\right)\left(e^{-\lambda_{23}T}\right)\left(e^{-\lambda_{123}T}\right) \\ &= \left(3e^{-2\lambda_{1/3}T} - 2e^{-3\lambda_{1/3}T}\right)\left[e^{-(3\lambda_{2/3}+\lambda_{3/3})T}\right] \end{aligned} \quad (2.13)$$

2.5.3 3oo3 System

Consider a 3oo3 system (i.e., a series system) in which all 3 components must be reliable to ensure the system reliability. The system reliability without CCF is given as

$$R_s(T) = R_1^3 = e^{-3\lambda_{1/3}T} \quad (2.14)$$

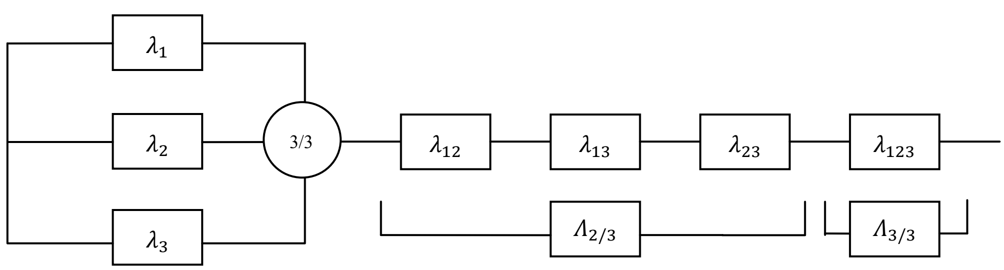


Figure 2.6: Cut sets of 3oo3 system with CCF

Taking into account four possible CCF events, all cut sets are shown above. The system reliability is given as

$$R_S'(T) = R_S(T)\left(e^{-\lambda_{12}T}\right)\left(e^{-\lambda_{13}T}\right)\left(e^{-\lambda_{23}T}\right)\left(e^{-\lambda_{123}T}\right)$$
$$= \left(e^{-3\lambda_{1/3}T}\right)\left(e^{-\left(3\lambda_{2/3}+\lambda_{3/3}\right)T}\right) = e^{-\left(3\lambda_{1/3}+3\lambda_{2/3}+\lambda_{3/3}\right)T} = \mathrm{e}^{-T\ \ 3} \quad (2.15)$$

2.5.4 Example

Suppose failure rates of a motor-operated valve (MOV) in the 'fail to open' (FO) mode were given as shown in the Table 2.1.

Table 2.1: Failure rates of a component

k	1	2	3
$\lambda_{k/3}$	0.04321	0.001543	0.00463

Substituting the above failure rates into Equation (2.10) to (2.15) and assuming the operation time of T = 100 (month), the system reliability of in various k-oo-n configurations can be calculated (Table 2.2). It is clear that system reliability decreases when CCF events are taken into account in the system reliability analysis. Of course, the decrease in the reliability depends on the failure rate of CCF events.

Table 2.2: Reliability of a k-out-of-3 system

System	System Reliability	
*k-out-of-*3	Without CCF	With CCF
1-*out-of-*3	3.93×10^{-2}	2.48×10^{-2}
2-*out-of-*3	5.25×10^{-4}	2.08×10^{-4}
3-*out-of-*3	2.35×10^{-6}	9.29×10^{-7}

2.6 Summary

This Section introduces the basic concepts and terminology associated with the modeling of common cause failures. The “General Multiple Failure Rate” (GMFR) model is described which is the basis of the CCF analysis in the current reliability literature. The GMFR model is utilized by the Nordic PSA Group for CCF analysis.

Section 3:
Data Mapping Methods

3.1 Introduction

CCF data for nuclear power plant safety system are scarce, such that text book methods of statistical estimation are not applicable to analyze failure rates. Therefore, conceptual methods were developed to assimilate whatever CCF data available from systems of various sizes in analyzing a particular system, also called the "target" system. This process is in general called data mapping, i.e., mapping source data from systems of all different group sizes to the target k-oo-n system (Mosleh *et al*., 1989, Vaurio, 2007).

The mapping down means the mapping of the source data from systems with CCG greater than n to a target k-oo-n system. The mapping up means the mapping of the source data from systems with CCG less than n to a target k-oo-n system.

The mapping is an important system to augment the data for increasing statistical confidence in the estimation of CCF rates. This Section describes the common mapping procedures.

3.2 Mapping Down of Data

The basic concept of mapping down the failure rates from a CCG 2 to the rate of CCG 1 is illustrated in Figure 3.1 (Vaurio, 2007). In case of CCG 2, CCF rates are related to the component rates as follows: $\Lambda_{2/2} = \lambda_{2/2}, \Lambda_{1/2} = 2\lambda_{1/2}$. These rates are now mapped to CCG of 1 by considering that a single failure can be caused by Poisson processes with rate $\lambda_{1/2}$ as well as $\lambda_{2/2}$, such that $\Lambda_{1/1} = \lambda_{1/1} = \lambda_{1/2} + \lambda_{2/2}$. In other words, the mapping down lead to the rates for CCG of 1 as

$$\Lambda_{1/1} = \frac{1}{\binom{2}{1}}\Lambda_{1/2} + \Lambda_{2/2} \qquad (2.16)$$

In general, this concept can be applied to mapping down the known CCF rates from CCG of n to the rates of CCG($n-1$), which lead to the following relations:

$$\frac{1}{\binom{n-1}{k}}\Lambda_{k/n-1} = \frac{1}{\binom{n}{k}}\Lambda_{k/n} + \frac{1}{\binom{n}{k+1}}\Lambda_{k+1/n} \tag{2.17}$$

$$\rightarrow \Lambda_{k/n-1} = \frac{n-k}{n}\Lambda_{k/n} + \frac{k+1}{n}\Lambda_{k+1/n} \tag{2.18}$$

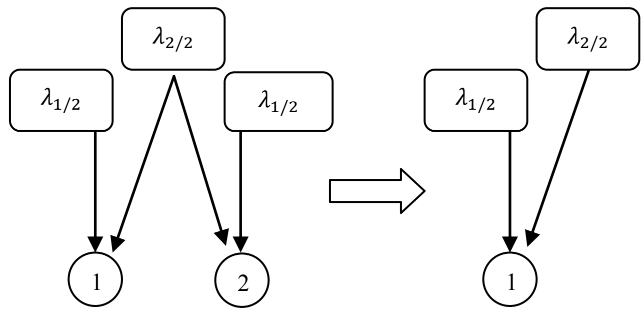


Figure 2.7: Mapping down CCF rates from CCG 2 to 1

The mapping down formula from CCG *n to* (*n-2*) can be written as

$$\frac{1}{\binom{n-2}{k}}\Lambda_{k/n-2} = \frac{1}{\binom{n}{k}}\Lambda_{k/n} + \frac{2}{\binom{n}{k+1}}\Lambda_{k+1/n} + \frac{1}{\binom{n}{k+2}}\Lambda_{k+2/n} \tag{2.19}$$

For example, using Equation (2.18) failure rates from CCG 4 to 3 can be mapped as

$$\Lambda_{3/3} = \frac{1}{4}\Lambda_{3/4} + \Lambda_{4/4} \tag{2.20}$$

$$\Lambda_{2/3} = \frac{1}{2}\Lambda_{2/4} + \frac{3}{4}\Lambda_{4/4} \tag{2.21}$$

$$\Lambda_{1/3} = \frac{3}{4}\Lambda_{1/4} + \frac{1}{2}\Lambda_{2/4} \tag{2.22}$$

Equations (3.3) and (3.4) are consistent with the mapping down formulas given by NUREG/CR-4780 listed in Table 2.3.

Table 2.3: Mapping down formulas for CCG up to 4 (NUREG-4780).

		Size of system mapping to		
		3	*2*	*1*
Size of system mapping from	*4*	$\Lambda_{1/3} = \frac{3}{4}\Lambda_{1/4} + \frac{1}{2}\Lambda_{2/4}$ $\Lambda_{2/3} = \frac{1}{2}\Lambda_{2/4} + \frac{3}{4}\Lambda_{3/4}$ $\Lambda_{3/3} = \frac{1}{4}\Lambda_{3/4} + \Lambda_{4/4}$	$\Lambda_{1/2} = \frac{1}{2}\Lambda_{1/4} + \frac{2}{3}\Lambda_{2/4} + \frac{1}{2}\Lambda_{3/4}$ $\Lambda_{2/2} = \frac{1}{6}\Lambda_{2/4} + \frac{1}{2}\Lambda_{3/4} + \Lambda_{4/4}$	$\Lambda_{1/1} = \frac{1}{4}\Lambda_{1/4} + \frac{1}{2}\Lambda_{2/4} + \frac{3}{4}\Lambda_{3/4} + \Lambda_{4/4}$
	3		$\Lambda_{1/2} = \frac{2}{3}\Lambda_{1/3} + \frac{2}{3}\Lambda_{2/3}$ $\Lambda_{2/2} = \frac{1}{3}\Lambda_{2/3} + \Lambda_{3/3}$	$\Lambda_{1/1} = \frac{1}{3}\Lambda_{1/3} + \frac{2}{3}\Lambda_{2/3} + \Lambda_{3/3}$
	2			$\Lambda_{1/1} = \frac{1}{2}\Lambda_{1/2} + \Lambda_{2/2}$

Kvam and Miller (2002) derived mapping down formulas that can be written in a fairly concise manner:

$$\Lambda_{k/m} = \sum_{r=1}^{n} C_{k,r}\Lambda_{r/n} \quad for\ (k = 1, \dots, m)$$
$$where \quad C_{k,r} = \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}} \quad (k \le r \le n - m + k) \qquad (2.23)$$

In order to verify this formula, take $n = 4$ and $m = 3$ to map failure rates from CCG 4 to 3. Consider $\Lambda_{2/3}$ as an example with $k = 2$ and $2 \le r \le 3$ from the above relation.

$$C_{2,2} = \frac{\binom{2}{2}\binom{4-2}{3-2}}{\binom{4}{3}} = \frac{1}{2} \quad (2.24)$$

$$C_{2,3} = \frac{\binom{3}{2}\binom{4-3}{3-3}}{\binom{4}{3}} = \frac{3}{4} \quad (2.25)$$

$$\Lambda_{2/3} = C_{2,2}\Lambda_{2/4} + C_{2,3}\Lambda_{3/4} = \frac{1}{2}\Lambda_{2/4} + \frac{3}{4}\Lambda_{3/4} \quad (2.26)$$

Equation (3.11) is the same as Equation (3.6).

3.3 Mapping Up of Data

3.3.1 NUREG-4780 Procedure

The mapping down is a rather deterministic procedure. It means that given a set of failure rates of a larger system, the estimation of rates for a smaller system is straightforward. However, estimation of failure rates by mapping up from a smaller system to a larger one is not deterministic, and extra assumptions have to be introduced in the estimation.

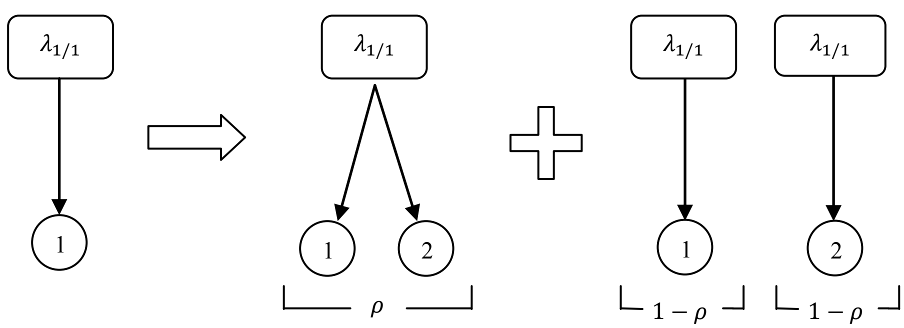


Figure 2.8: Mapping up from CCG 1 to 2.

An illustration of mapping from size 1 to 2 is shown in Figure 3.2. Here the ideas is that the HPP with failure rate $\lambda_{1/1}$ can cause a CCF event in CCG 2 with a probability ρ, and it can also cause single failures with probability $(1-\rho)$. Based on this thinking, the following mapping up equations are obtained:

$$\Lambda_{2/2} = \rho\Lambda_{1/1} \tag{2.27}$$

$$\Lambda_{1/2} = 2(1-\rho)\Lambda_{1/1} \tag{2.28}$$

It can be seen that an additional parameter, ρ, is needed to perform the mapping up procedure. Table 2.4 provides the mapping up formulas for CCG of 1 to 4.

Table 2.4: Formulas for mapping up failure rates for CCG 1 to 4 (NUREG-4780).

		Size of system mapping to		
		2	*3*	*4*
Size of system mapping from	*1*	$\Lambda_{1/2} = 2(1-\rho)\Lambda_{1/1}$ $\Lambda_{2/2} = \rho\Lambda_{1/1}$	$\Lambda_{1/3} = 3(1-\rho)^2\Lambda_{1/1}$ $\Lambda_{2/3} = 3\rho(1-\rho)\Lambda_{1/1}$ $\Lambda_{3/3} = \rho^2\Lambda_{1/1}$	$\Lambda_{1/4} = 4(1-\rho)^3\Lambda_{1/1}$ $\Lambda_{2/4} = 6\rho(1-\rho)^2\Lambda_{1/1}$ $\Lambda_{3/4} = 4\rho^2(1-\rho)\Lambda_{1/1}$ $\Lambda_{4/4} = \rho^3\Lambda_{1/1}$
	2		$\Lambda_{1/3} = \frac{3}{2}(1-\rho)\Lambda_{1/2}$ $\Lambda_{2/3} = \rho\Lambda_{1/2} + (1-\rho)\Lambda_{2/2}$ $\Lambda_{3/3} = \rho\Lambda_{2/2}$	$\Lambda_{1/4} = 2(1-\rho)^2\Lambda_{1/2}$ $\Lambda_{2/4} = \frac{5}{2}\rho(1-\rho)\Lambda_{1/2} + (1-\rho)^2\Lambda_{2/2}$ $\Lambda_{3/4} = \rho^2\Lambda_{1/2} + 2\rho(1-\rho)\Lambda_{2/2}$ $\Lambda_{4/4} = \rho^2\Lambda_{2/2}$
	3			$\Lambda_{1/4} = \frac{4}{3}(1-\rho)\Lambda_{1/3}$ $\Lambda_{2/4} = \rho\Lambda_{1/3} + (1-\rho)\Lambda_{2/3}$ $\Lambda_{3/4} = \rho\Lambda_{2/3} + (1-\rho)\Lambda_{3/3}$ $\Lambda_{4/4} = \rho\Lambda_{3/3}$

3.3.2 Method of Vaurio (2007)

Since it is unreasonable to assume a single parameter ρ for all the CCG sizes and CCF events, Vaurio (2007) proposed a set of different parameters for the mapping up from n to $n+1$:

$$\Lambda_{n+1/n+1} = \eta_{n,n+1}\Lambda_{n/n} \tag{2.29}$$

Substituting successively $n = 2, 3, 4, \ldots$ and $= n-1, n-2, \ldots$, equation (3.14) leads to the following mapping up formulas:

$$\Lambda_{2/2} = \eta_{1,2}\Lambda_{1/1} \tag{2.30}$$

$$\Lambda_{1/2} = 2(1-\eta_{1,2})\Lambda_{1/1} \tag{2.31}$$

$$\Lambda_{3/3} = \eta_{2,3}\Lambda_{2/2} = \eta_{2,3}\eta_{1,2}\Lambda_{1/1} \tag{2.32}$$

$$\Lambda_{2/3} = 3(1-\eta_{2,3})\Lambda_{2/2} = 3(1-\eta_{2,3})\eta_{1,2}\Lambda_{1/1} \tag{2.33}$$

$$\Lambda_{1/3} = \frac{3}{2}\Lambda_{1/2} - 3(1-\eta_{2,3})\Lambda_{2/2} = 3[1-(2-\eta_{2,3})\eta_{1,2}]\Lambda_{1/1} \tag{2.34}$$

It can be seen that all the total failure rates $\Lambda_{k/n}$ can be represented by $\Lambda_{1/1}$ multiplied by a product of the extra parameters. If we set all the parameters as the same and express the traditional mapping up results listed in Table 2.4 in terms of $\Lambda_{1/1}$, we can see that the two mapping up procedures generate the same results.

The value of this parameter $\eta_{n,n+1}$ is not known a priori. Vaurio (2007) proposed the following formula that uses alpha factors.

$$\eta_{n,n+1} = \frac{(n+1)\alpha_{n+1,n+1}}{(n+\alpha_{2,n+1}+\cdots+\alpha_{n+1,n+1})\alpha_{n,n}} \tag{2.35}$$

Compared to the traditional mapping up process which utilizes a single extra parameter ρ, the introduction of $\eta_{n,n+1}$ is making more sense and does not completely rely on expert judgments.

3.4 Summary

Since the CCF data are sparse, probabilistic concepts are developed to borrow the data from different CCG sizes for the estimation of CCF in a target system. This Section provides an overview of the up and down mapping of the data and discusses underlying concepts.

The data mapping methods will be utilized in the Bayesian analysis and the case studies presented in next Sections.

.

Section 4:
Statistical Estimation of Failure Rates

4.1 Introduction

After collecting the CCF data, next step is to estimate failure rates. The maximum likelihood (ML) method is a standard estimation method. Since CCF dataset are sparse, ML estimates are not considered robust. Also the confidence interval associated with the rates tends to be fairly wide due to lack of data. Therefore, the development of the Bayesian estimation method has been actively pursued by PSA experts.

In the Bayesian method, a prior distribution is assigned to the failure rate, which serves the purpose of modeling parameter (or epistemic) uncertainty associated the failure rate. Now the problem is shifted to the estimation of the prior distribution. In the context of Poisson process model, the gamma distribution is a conjugate prior to the failure rate, which simplifies considerably the mathematical steps associated with the Bayesian updating of failure rate. The remaining problem is the estimation of parameters of the gamma prior. For this purpose, Empirical Bayes (EB) has been popularly used in the reliability literature. The EB method also allows to combine the data from different sub-systems and plants. The Bayesian modeling of the failure rate using gamma prior and applying the EB method for estimation is referred to as Parametric Empirical Bayes (PREB) in the CCF literature. PREB has been adopted by Nordic countries for the estimation of CCF rates.

This Section describes the maximum likelihood (MLE) and the Parametric Empirical Bayes (PREB) methods for the estimation of CCF rates. A numerical example is presented to compare the results obtained by these two methods.

4.2 Maximum Likelihood Estimation (MLE)

The maximum likelihood is the simplest method for the estimation of the failure rate of a HPP model. If N_i failures are observed in the duration T_i, the failure rate and the associated standard error are estimated as (Crowder *et al.* 1991):

$$\hat{\lambda}_i = \frac{N_i}{T_i} \tag{4.1}$$

$$\sigma(\hat{\lambda}_i) = \sqrt{\frac{N_i}{{T_i}^2}} \tag{4.2}$$

Using the CCF data from an individual system or the mapped data, the MLE of failure rate is fairly easy to compute.

4.3 Bayesian Method

4.3.1 Basic Concepts

In the Bayesian Method, a prior distribution, $\pi(\lambda)$, is assigned to the failure rate λ, which is usually determined based on the past experience and expert judgment. The conjugate prior, the gamma prior distribution, is a common choice due to mathematical simplicity and its flexibility to fit various types of data (Carlin and Louis, 2000)

$$\pi(\lambda_i) = \frac{\beta^{\alpha} {\lambda_i}^{\alpha-1} e^{-\beta\lambda_i}}{\Gamma(\alpha)} \quad , \qquad \lambda_i, \alpha, \beta > 0 \tag{4.3}$$

The mean and variance of the prior are

$$M = \alpha/\beta \;\text{ and }\; V = \alpha/\beta^2 \tag{4.4}$$

Poisson likelihood for the failure events is given as

$$L[N_i|\lambda_i, T_i] = \frac{e^{-\lambda_i T_i}(\lambda_i T_i)^{N_i}}{N_i!} \tag{4.5}$$

The posterior of the failure rate is also a gamma distribution:

$$p(\lambda_i|N_i, T_i) = \frac{(\beta + T_i)^{(\alpha+N_i)} {\lambda_i}^{\alpha+N_i-1} e^{-(\beta+T_i)\lambda_i}}{\Gamma(\alpha + N_i)} \tag{4.6}$$

The mean and variance of the gamma posterior are given as

$$\widehat{M}_i = (\alpha + N_i)/(\beta + T_i) \;\text{ and }\; \hat{V}_i = (\alpha + N_i)/(\beta + T_i)^2 \tag{4.7}$$

Typically the mean the of the posterior is reported as the estimate of the failure rate.

4.3.2 Empirical Bayes Method

Empirical Bayes (EB) is a method for the estimation of parameters of a prior distribution used in the Bayesian analysis. In this Section EB method proposed by Vaurio (2007) is described, since it is adopted by the Nordic PSA group.

The basic idea is that the failure data of n components are generated by a Poisson process and the failure rate for each component is a realization from a single Gamma prior with hyper-parameters α and β. The EB method is applied to estimate the parameters α and β using the past event data. Then the distribution for a particular plant is obtained from the updated posterior of the distribution, as described in the previous Section. A novel feature is that in the pooling of the data collected from different systems, proper weights are assigned. The steps of this method are given in Table 4.1.

Table 4.1: EB method for estimating the gamma prior from data

$T = \sum_{i=1}^{n} T_i$	*(4.8)*
$T^* = T - max\ (T_i)$	*(4.9)*
$m = \sum_{i=1}^{n} w_i N_i / T_i$	*(4.10)*
$S = R \sum_{i=1}^{n} w_i (N_i/T_i - m)^2$ where $R = 1/(1 - \sum_{i=1}^{n} {w_i}^2)$	*(4.11)*
$M_0 = m$	*(4.12)*
$V_0 = S + M_0/T^*$	*(4.13)*
$u_i = T_i/(T_i + M_0/V_0)$	*(4.14)*

$$w_i = u_i / \sum_{j=1}^{n} u_j$$	*(4.15)*
$$\alpha = M^2/V \text{ and } \beta = M/V$$	*(4.16)*

Note that the initial value of the weight w_i is $1/n$. Iterating through Equation (4.10) to (4.15) until M_0, V_0 and w_i converge, the estimates of the posterior mean and variance are obtained.

4.4 Numerical Example

The data, (N_i,T_i), regarding the number of failures and time of exposure for a group of 8 components are taken from Vaurio (1992) and summarized in Table 4.2.

Table 4.2: MLE and EB estimates of failure rates

i	N_i	T_i (10^6h)	MLE		Vaurio's EB		
			Mean	S.D.	w_i	Mean	S.D.
1	31	236.9020	0.1309	0.0235	0.1539	0.1343	0.0238
2	157	115.9440	1.3541	0.1081	0.1534	1.3532	0.1077
3	30	36.8120	0.8150	0.1488	0.1513	0.8229	0.1480
4	13	7.5970	1.7112	0.4746	0.1405	1.6664	0.4469
5	7	5.4660	1.2806	0.4840	0.1358	1.2723	0.4525
6	7	1.6890	4.1445	1.5665	0.1070	3.2439	1.1536
7	0	1.1230	0.0000	0.0000	0.0926	0.4846	0.5089
8	0	0.5520	0.0000	0.0000	0.0655	0.6974	0.7323

MLE estimates are shown for all the components. EB method given in Table 4.1 leads to the following estimates: $\alpha = 0.9070$ and $\beta = 0.7485$. Then posterior mean and standard deviation (S.D.) were computed for each component.

The mean and standard deviation of failure rates of a selected set of components are compared in Figures 4.1.

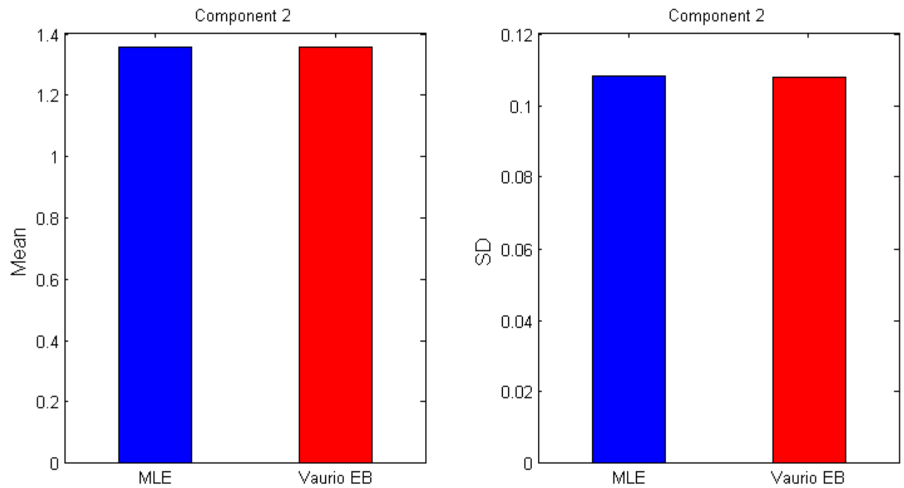


Figure 4.1 (a): Component 2

As seen from Table 4.2 that component 2 has large number of failures in a relatively large exposure time. Therefore, the MLE and EB estimates of the mean and SD are almost the same.

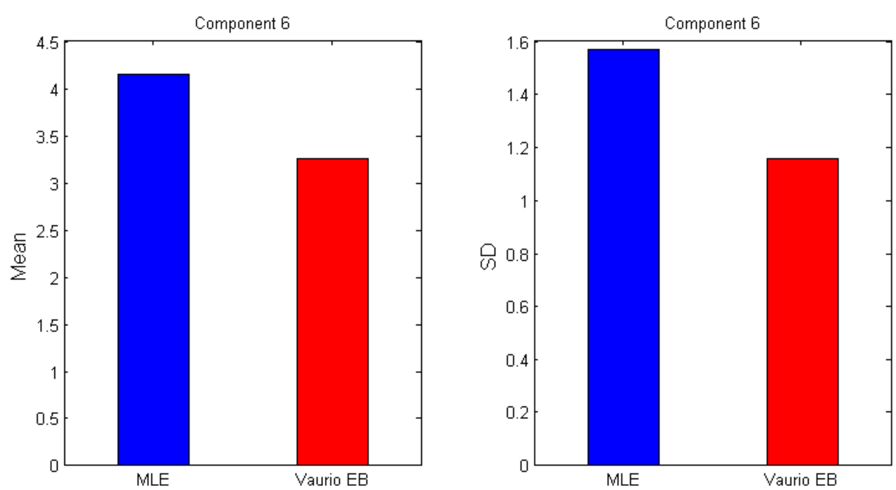


Figure 4.1 (b): Component 6

Component 6 has a relatively small number of failures and the exposure time. In this case, the EB estimate of mean and SD are smaller than the corresponding ML estimates.

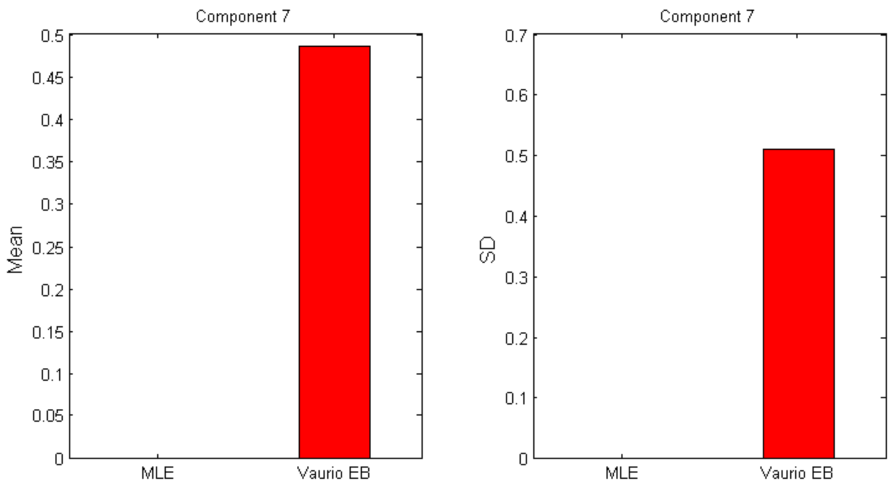


Figure 4.1 (c): Component 7

Figure 4.1: Comparison of mean and SD of failure rates (MLE and EB methods)

In case of component 7, no failures have been recorded. Therefore, MLE point estimates are zero. Nevertheless, EB provides estimates of mean and SD. This is a useful feature of EB method.

Prior and posterior distributions of the failure rate for selected components are shown in Figures 4.2 to 4.4. As expected, posterior distributions are narrower (less uncertain) than the prior distributions.

Figure 4.2: Distribution of the failure rate: component 1

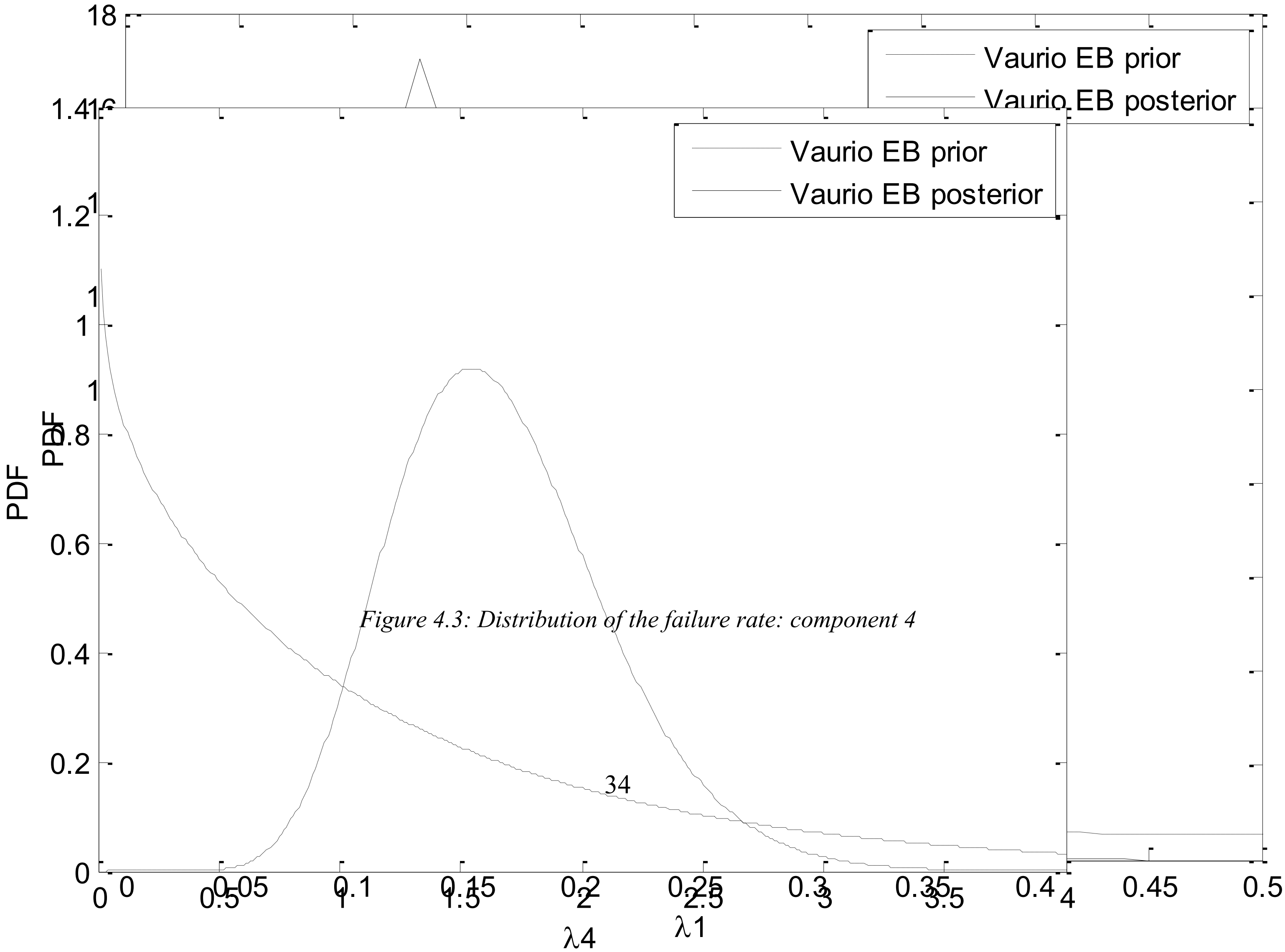


Figure 4.3: Distribution of the failure rate: component 4

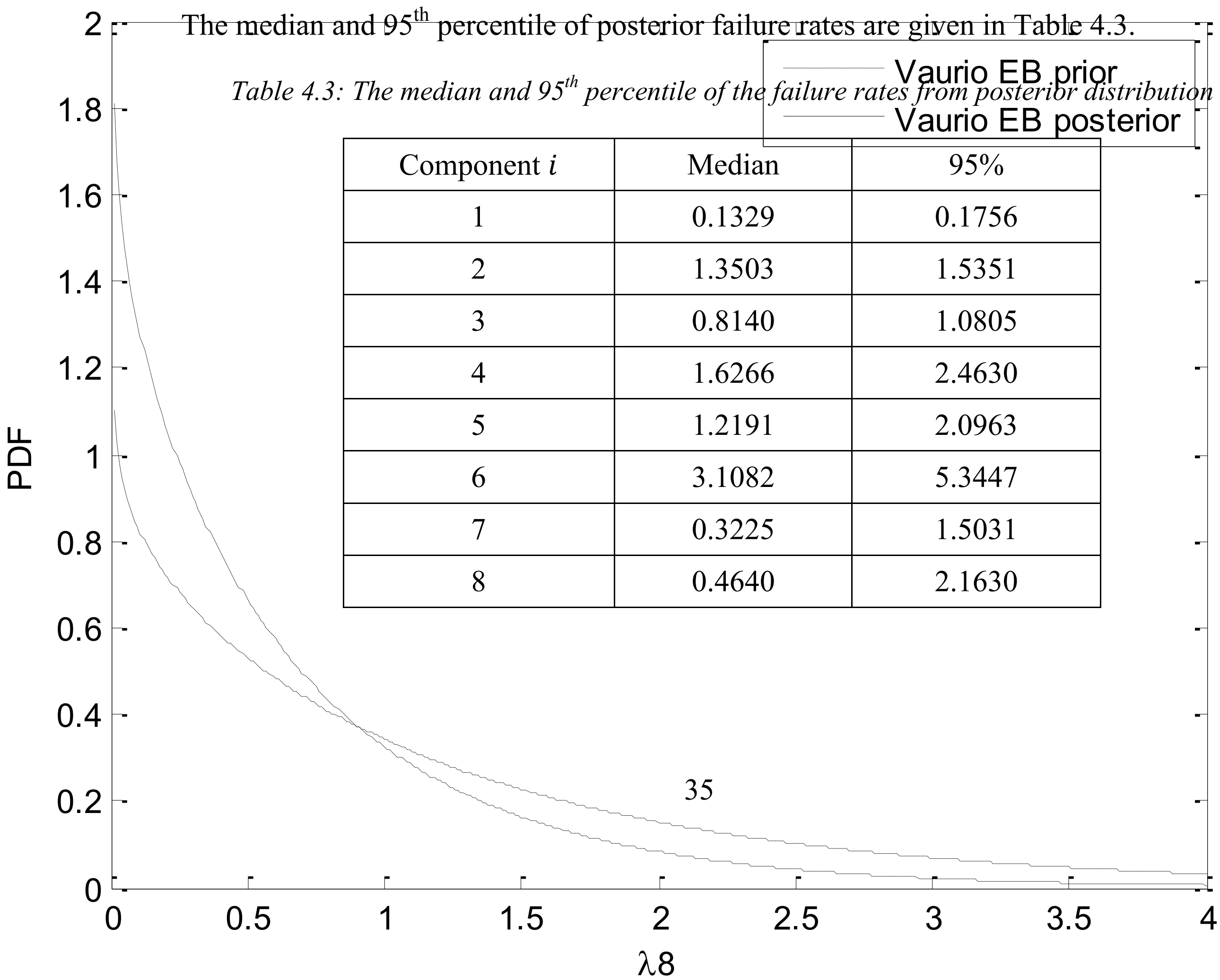


Figure 4.4: Distribution of the failure rate: component 8

The median and 95th percentile of posterior failure rates are given in Table 4.3.

Table 4.3: The median and 95th percentile of the failure rates from posterior distribution

Component i	Median	95%
1	0.1329	0.1756
2	1.3503	1.5351
3	0.8140	1.0805
4	1.6266	2.4630
5	1.2191	2.0963
6	3.1082	5.3447
7	0.3225	1.5031
8	0.4640	2.1630

4.5 Summary

In this chapter, MLE and EB methods are described for the estimation of failure rates. In case of a small data set, the ML estimate tends to have larger uncertainty. The EB method is useful in pooling the data and also estimating failure rates in cases of zero observations of failure.

Section 5:
Case Studies

5.1 Introduction

This Section presents two comprehensive case studies of analyzing the CCF data for the motor operated valves (MOV). The purpose is to illustrate the application of data mapping and statistical estimation methods (MLE and EB) in a practical setting.

The key objective is to perform statistical analysis suing the MLE and EB and method before and after the data mapping. This detailed analysis allows to understand the impact of assumptions used in each analysis. The target system is a parallel system of 4 MOVs.

5.2 Example 1: Motor Operated Valves

The first data set consists of CCF data from MOV CCG of 2, 4, 8 and 16, as shown in Table 5.1. These data were collected over an 18 year period. The MOV failure mode is "failure to open" (FO).

In the ICDE database component states are defined as complete failure (C), degraded (D), incipient (I) and working (W). In a formal analysis, impact vectors are assigned depending on the state of the system. For sake of simplicity (and lack of data), no distinction is made among the states C, D and I, and they are treated as the failure.

Table 5.1: CCF Data: case study 1

System size	No. of systems	Total Time	No. of failures					
n	m	m*18*12 (month)	$1/n$	$2/n$	$3/n$	$4/n$	$5/n$	$6/n$
2	49	10584	36	1	0	0	0	0
4	17	3672	18	2	10	1	0	0
8	8	1728	6	1	0	0	0	0
16	5	1080	13	1	0	0	0	1

5.2.1 MLE without Data Mapping

In this case only the data for CCG collected from a population of 17 systems over an 18 year period is analyzed using the MLE method. There is no data mapping considered here. The mean and SD of failure rates $\Lambda_{k/4}$, and corresponding α-factors are given in Table 5.2.

Table 5.2: MLE results: case study 1

Multiplicity	k	1	2	3	4
Number of failures	$N_{k/4}$	18	2	10	1
Mean of failure rate (/month)	$\Lambda_{k/4}$	4.90×10^{-3}	5.45×10^{-4}	2.72×10^{-3}	2.72×10^{-4}
Standard deviation (/month)	$\sigma(\Lambda_{k/4})$	1.16×10^{-3}	3.85×10^{-4}	8.61×10^{-4}	2.72×10^{-4}
α factor	$\alpha_{k/4}$	5.81×10^{-1}	6.45×10^{-2}	3.23×10^{-1}	3.23×10^{-2}

Using the relation of $\Lambda_{k/4} = \binom{4}{k}\lambda_{k/4}$, the mean and variance of the component specific failure rates can also be calculated. Note that α factors are calculated from Equations (2.4) and (2.5).

5.2.2 MLE with Data Mapping

The CCF data of CCG 8 and 16 are mapped down to CCG 4 and data from CCG 2 is mapped to CCG 4. In the mapping up the data, parameter ρ is assumed as 0.2 according to NUREG-4780. The results of data mapping are given in Table 5.3.

In MLE analysis, all the mapped number of failures of a particular multiplicity koon are summed, as well as the corresponding exposure times. The total operation time is 17064 month. The MLE analysis of mapped data leads to the results shown in Table 5.4. The comparison of results obtained with and without mapping is given in the last Section of the report.

Table 5.3: CCF data mapped to CCG of 4

System size		Number of failures $N_{k/n}$					
n		$1/n$	$2/n$	$3/n$	$4/n$	$5/n$	$6/n$
2	original	36	1	0	0	0	0
	mapped	36	0.6400	0.3200	0.0400	0	0
4	original	18	2	10	1	0	0
	mapped	18	2	10	1	0	0
8	original	6	1	0	0	0	0
	mapped	3.5714	0.2143	0	0	0	0
16	original	13	1	0	0	0	1
	mapped	4.0456	0.4209	0.1099	0.0082	0	0
sum	mapped	61.6170	3.2752	10.4299	1.0482	For HPP	

Table 5.4: MLE results with data mapping: case study 1

Multiplicity	k	1	2	3	4
Number of failures	$N_{k/4}$	61.6170	3.2752	10.4299	1.0482
Mean of failure rate (/month)	$\Lambda_{k/4}$	3.61×10^{-3}	1.92×10^{-4}	6.11×10^{-4}	6.14×10^{-5}
Standard deviation (/month)	$\sigma(\Lambda_{k/4})$	4.60×10^{-4}	1.06×10^{-4}	1.89×10^{-4}	6.00×10^{-5}
α factor	$\alpha_{k/4}$	8.07×10^{-1}	4.29×10^{-2}	1.37×10^{-1}	1.37×10^{-2}

5.2.3 Empirical Bayes (EB) without Data Mapping

EB method was applied to CCF dart for CCG of 4. The parameters of the gamma prior were estimated as $\alpha = 0.9367$ and $\beta = 434.4733$. The posterior mean and SD of failure rates are given in Table 5.5. The prior and posterior distributions of the system level failure rates are plotted in Figure 5.1.

Table 5.5: Posterior Mean and SD of failure rate without data mapping

Multiplicity	k	1	2	3	4
No. of failures	$N_{k/4}$	18	2	10	1
Mean of failure rate (/month)	$\Lambda_{k/4}$	4.61×10^{-3}	7.15×10^{-4}	2.66×10^{-3}	4.72×10^{-4}
Standard deviation (/month)	$\sigma(\Lambda_{k/4})$	1.06×10^{-3}	4.17×10^{-4}	8.05×10^{-4}	3.39×10^{-4}
α factor	$\alpha_{k/4}$	5.45×10^{-1}	8.45×10^{-2}	3.15×10^{-1}	5.57×10^{-2}

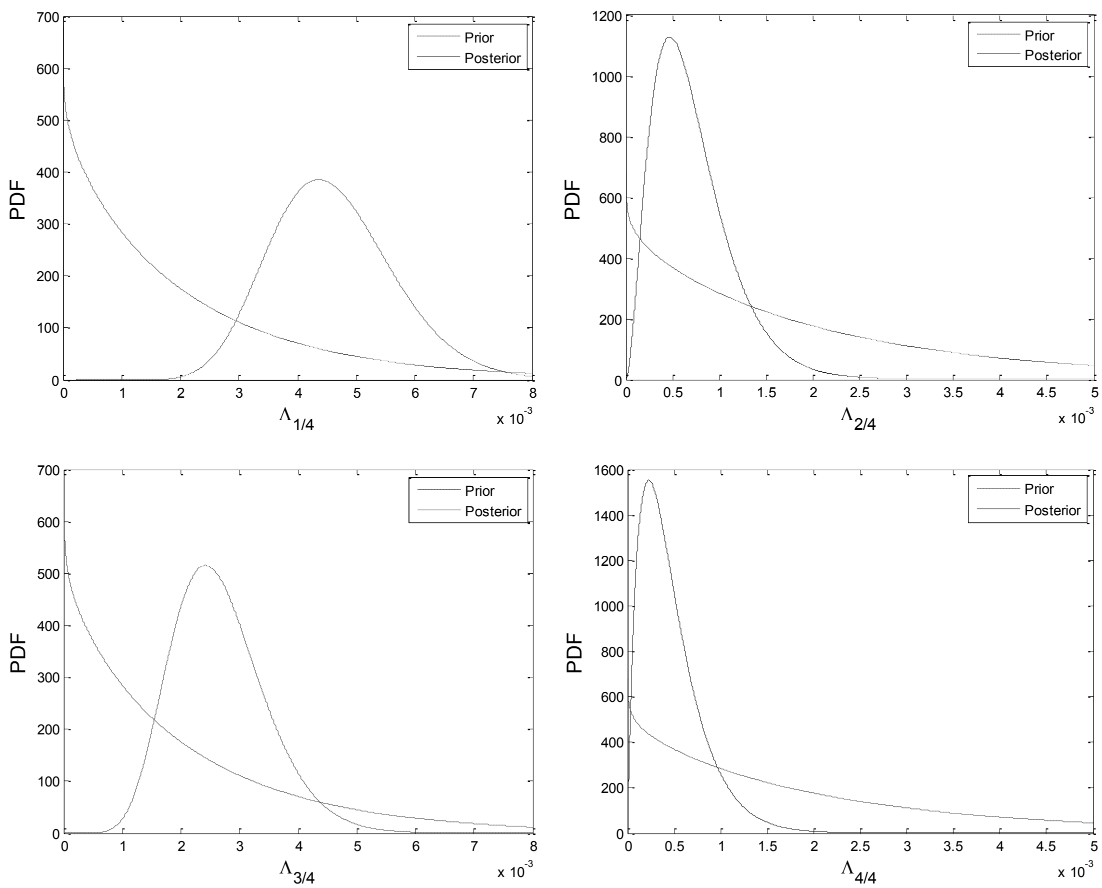


Figure 5.1: Failure rate distribution without data mapping: case study 1

5.2.4 EB with Data Mapping (Multiple Priors)

In this case mapped data given in Table 5.3 is used in EB analysis. A prior distribution is assigned to each failure rate $\Lambda_{k/4}, k = 1,..,4$. This way, there are 4 gamma priors and 4 sets of distribution parameters are estimated.

Table 5.6: Number and exposure times of 1oo4 failures after data mapping

System size n	$N_{1/4}$	T_n (month)
2	36	10584
4	18	3672
8	3.5714	1728
16	4.0456	1080

As an example, 1oo4 failure mapped data given in Table 5.3 is analyzed. The number of failures after data mapping and the corresponding exposure time are given in Table 5.6. The EB analysis leads to the posterior mean of $\Lambda_{1/4}$ as 4.49×10^{-3} failures per month.

Table 5.7: Posterior Mean and SD of failure rates with data mapping

Multiplicity	k	1	2	3	4
Mean of failure rate (/month)	$\Lambda_{k/4}$	4.49×10^{-3}	4.58×10^{-4}	2.55×10^{-3}	2.26×10^{-4}
Standard deviation (/month)	$\sigma(\Lambda_{k/4})$	8.97×10^{-4}	2.72×10^{-4}	7.93×10^{-4}	1.93×10^{-4}
α factor	$\alpha_{k/4}$	5.81×10^{-1}	5.94×10^{-2}	3.30×10^{-1}	2.93×10^{-2}

Repeating the above procedure, results were obtained remaining multiplicities as shown in Table 5.7. Parameters of the gamma prior are given in Table 5.8.

Table 5.8: Parameters of Priors (mapped data)

Failure rate	$\Lambda_{1/4}$	$\Lambda_{2/4}$	$\Lambda_{3/4}$	$\Lambda_{4/4}$
α	7.0049	0.8436	0.3118	0.3799
β	1902.2505	2534.1352	377.2051	2429.8139

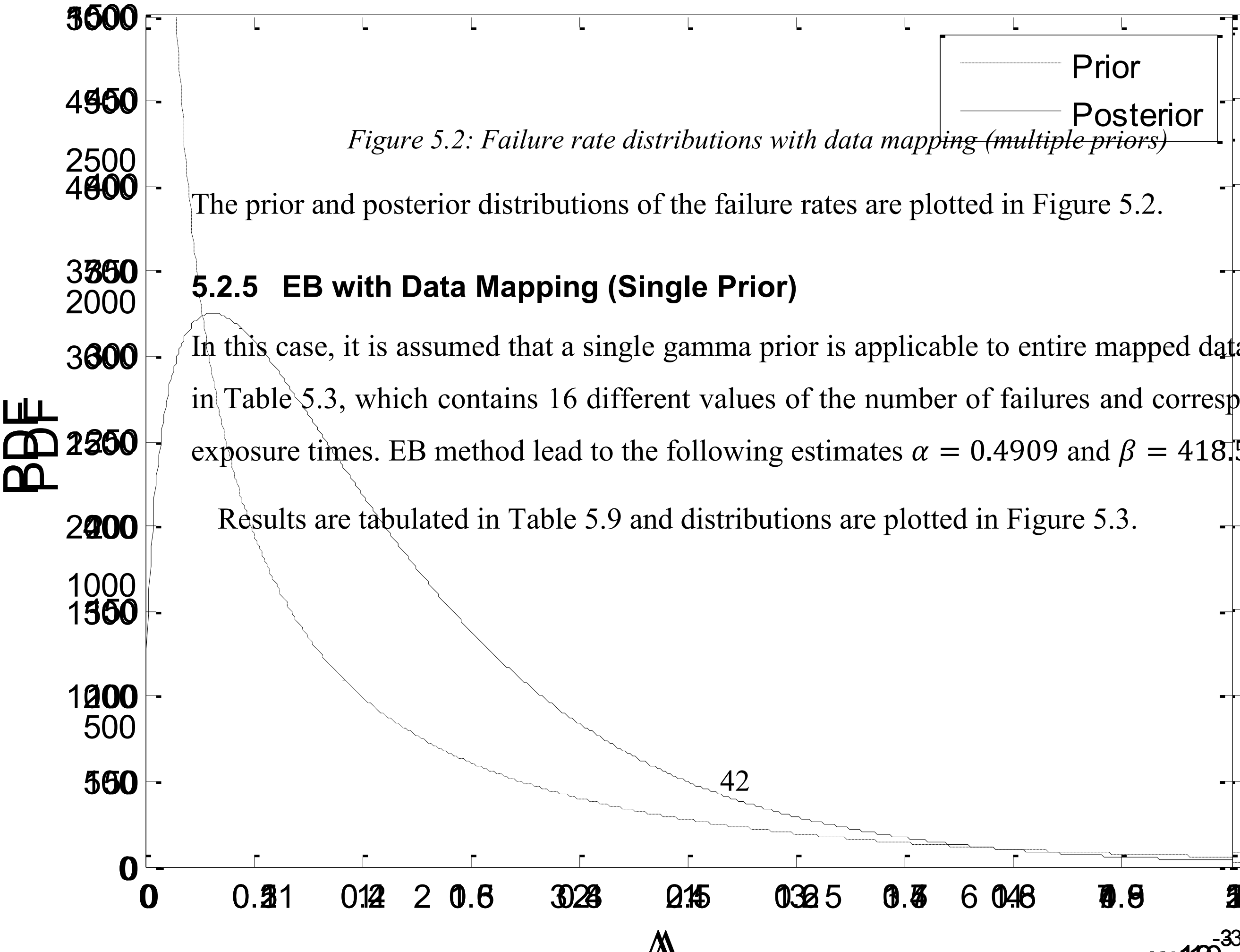


Figure 5.2: Failure rate distributions with data mapping (multiple priors)

The prior and posterior distributions of the failure rates are plotted in Figure 5.2.

5.2.5 EB with Data Mapping (Single Prior)

In this case, it is assumed that a single gamma prior is applicable to entire mapped data given in Table 5.3, which contains 16 different values of the number of failures and corresponding exposure times. EB method lead to the following estimates $\alpha = 0.4909$ and $\beta = 418.5876$.

Results are tabulated in Table 5.9 and distributions are plotted in Figure 5.3.

Table 5.9: Posterior mean and of failure rates with data mapping

Multiplicity	k	1	2	3	4
No. of failures	$N_{k/4}$	18	2	10	1
Mean of failure rate (/month)	$\Lambda_{k/4}$	4.52×10^{-3}	6.09×10^{-4}	2.56×10^{-3}	3.64×10^{-4}
Standard deviation (/month)	$\sigma(\Lambda_{k/4})$	1.05×10^{-3}	3.86×10^{-4}	7.92×10^{-4}	2.98×10^{-4}
α factor	$\alpha_{k/4}$	5.61×10^{-1}	7.56×10^{-2}	3.18×10^{-1}	4.52×10^{-2}

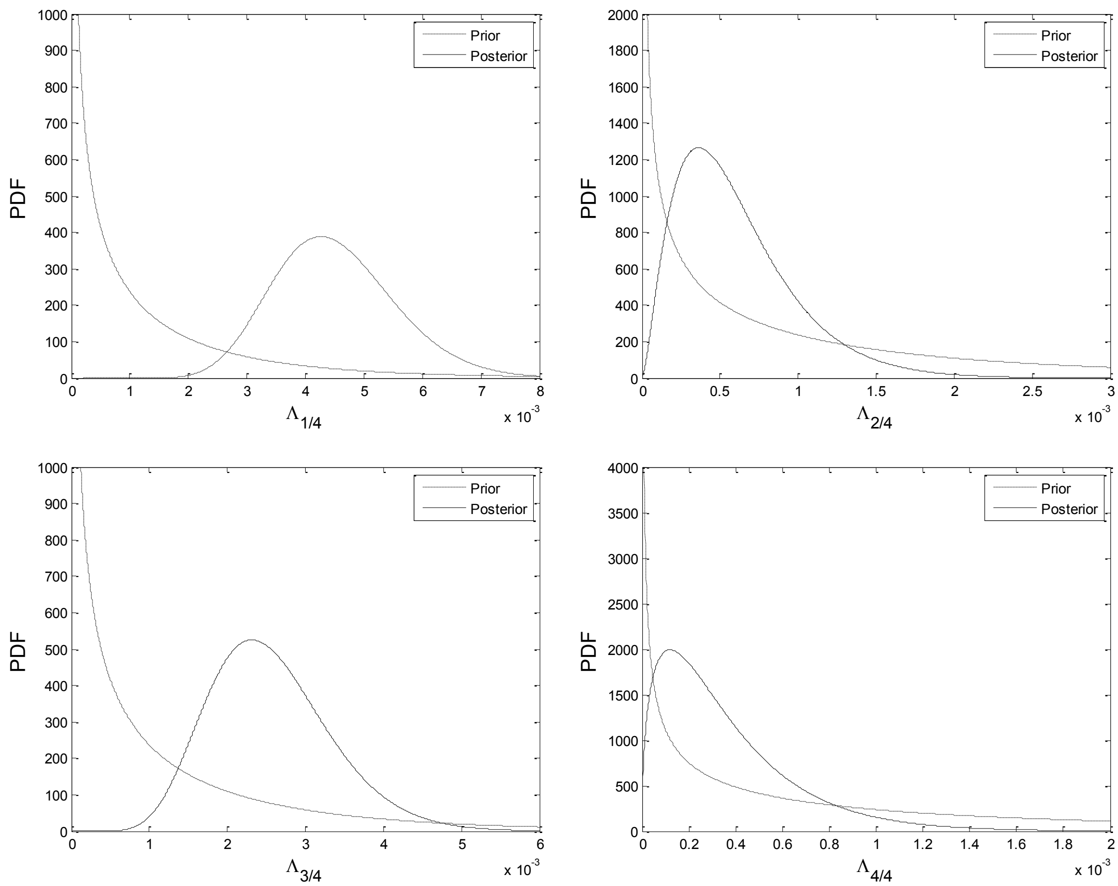


Figure 5.3: Failure rate distributions with data mapping (single prior)

5.2.6 Comparison of Results: Case Study 1

In order to understand the effect assumptions associated with 5 estimation methods, the mean of the failure rates are compared in Table 5.10 and graphically shown in Figure 5.4.

Table 5.10: Mean failure rates (per month) by different methods: case study 1

No.	Method	$\Lambda_{1/4}$	$\Lambda_{2/4}$	$\Lambda_{3/4}$	$\Lambda_{4/4}$
M1	MLE without data mapping	4.90×10^{-3}	5.45×10^{-4}	2.72×10^{-3}	2.72×10^{-4}
M2	MLE with data mapping	3.61×10^{-3}	1.92×10^{-4}	6.11×10^{-4}	6.14×10^{-5}
M3	EB without data mapping	4.61×10^{-3}	7.15×10^{-4}	2.66×10^{-3}	4.72×10^{-4}
M4	EB with data mapping (multiple priors)	4.49×10^{-3}	4.58×10^{-4}	2.55×10^{-3}	2.26×10^{-4}
M5	EB with data mapping (single prior)	4.52×10^{-3}	6.09×10^{-4}	2.56×10^{-3}	3.64×10^{-4}

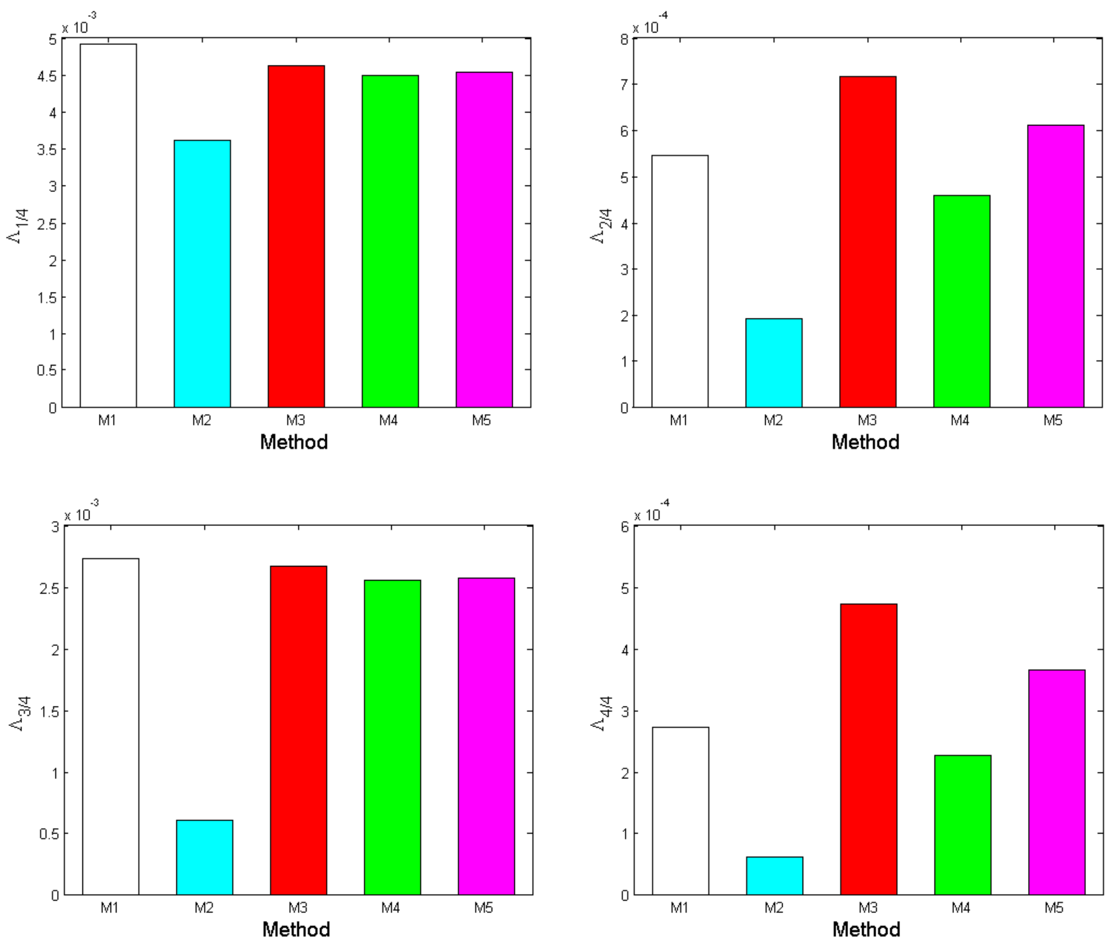
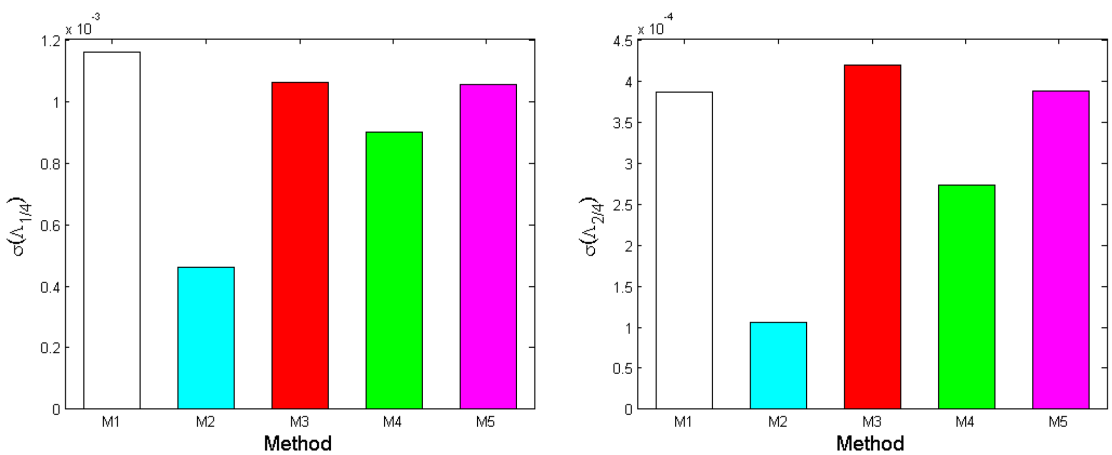


Figure 5.4: Comparison of mean failure rates: case study 1

- MLE with data mapping (M2) leads to lower values of mean failure rate as compared to that without data mapping (M1).
- EB with data mapping and single prior (M5) leads to higher mean rate than MMLE with data mapping (m2).

Table 5.11: SD of failure rates by different methods: case study 1

No.	Method	SD($\Lambda_{1/4}$)	SD($\Lambda_{2/4}$)	SD($\Lambda_{3/4}$)	SD($\Lambda_{4/4}$)
M1	MLE without data mapping	1.16×10^{-3}	3.85×10^{-4}	8.61×10^{-4}	2.72×10^{-4}
M2	MLE with data mapping	4.60×10^{-4}	1.06×10^{-4}	1.89×10^{-4}	6.00×10^{-5}
M3	EB without data mapping	1.06×10^{-3}	4.17×10^{-4}	8.05×10^{-4}	3.39×10^{-4}
M4	EB with data mapping (multiple priors)	8.97×10^{-4}	2.72×10^{-4}	7.93×10^{-4}	1.93×10^{-4}
M5	EB with data mapping (single prior)	1.05×10^{-3}	3.86×10^{-4}	7.92×10^{-4}	2.98×10^{-4}



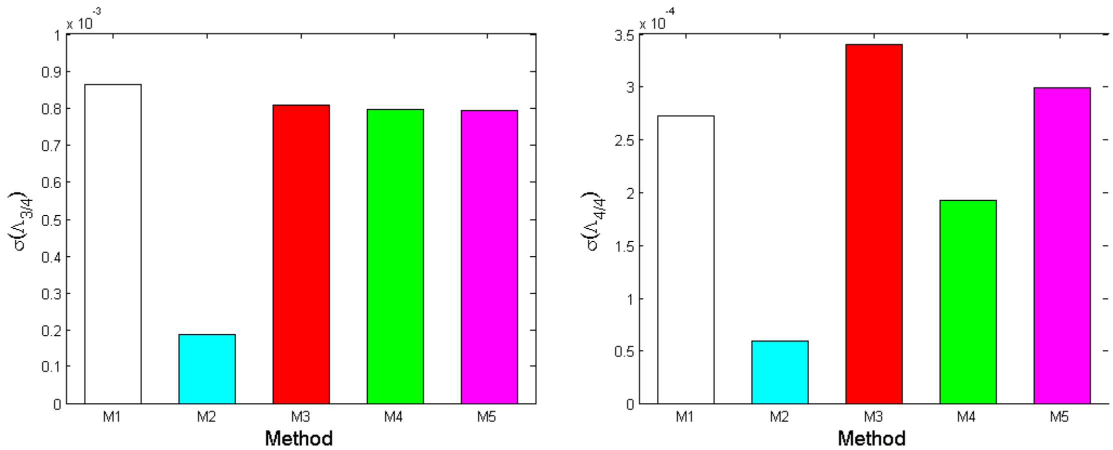


Figure 5.5: Comparison of standard deviations of failure rates by different methods

- SD of failure rate obtained by EB methods (M3 – M5) is fairly close.
- SD of EB method is slightly smaller than than that of MLE without data mapping (M1).

Table 5.12: Alpha factors by different method: case study 1

No.	Method	$\alpha_{1/4}$	$\alpha_{2/4}$	$\alpha_{3/4}$	$\alpha_{4/4}$
M1	MLE without data mapping	5.81×10^{-1}	6.45×10^{-2}	3.23×10^{-1}	3.23×10^{-2}
M2	MLE with data mapping	8.07×10^{-1}	4.29×10^{-2}	1.37×10^{-1}	1.37×10^{-2}
M3	EB without data mapping	5.45×10^{-1}	8.45×10^{-2}	3.15×10^{-1}	5.57×10^{-2}
M4	EB with data mapping (multiple priors)	5.81×10^{-1}	5.94×10^{-2}	3.30×10^{-1}	2.93×10^{-2}
M5	EB with data mapping (single prior)	5.61×10^{-1}	7.56×10^{-2}	3.18×10^{-1}	4.52×10^{-2}
Empirical	NUREG/CR-5497	9.69×10^{-1}	6.50×10^{-3}	2.50×10^{-3}	2.22×10^{-2}

In Table 5.12, alpha factors are compared along with the values given in NUREG/CR-5497 for 4- HPCI/RCIC injection MOV. Graphical comparison is shown in Figure 5.6.

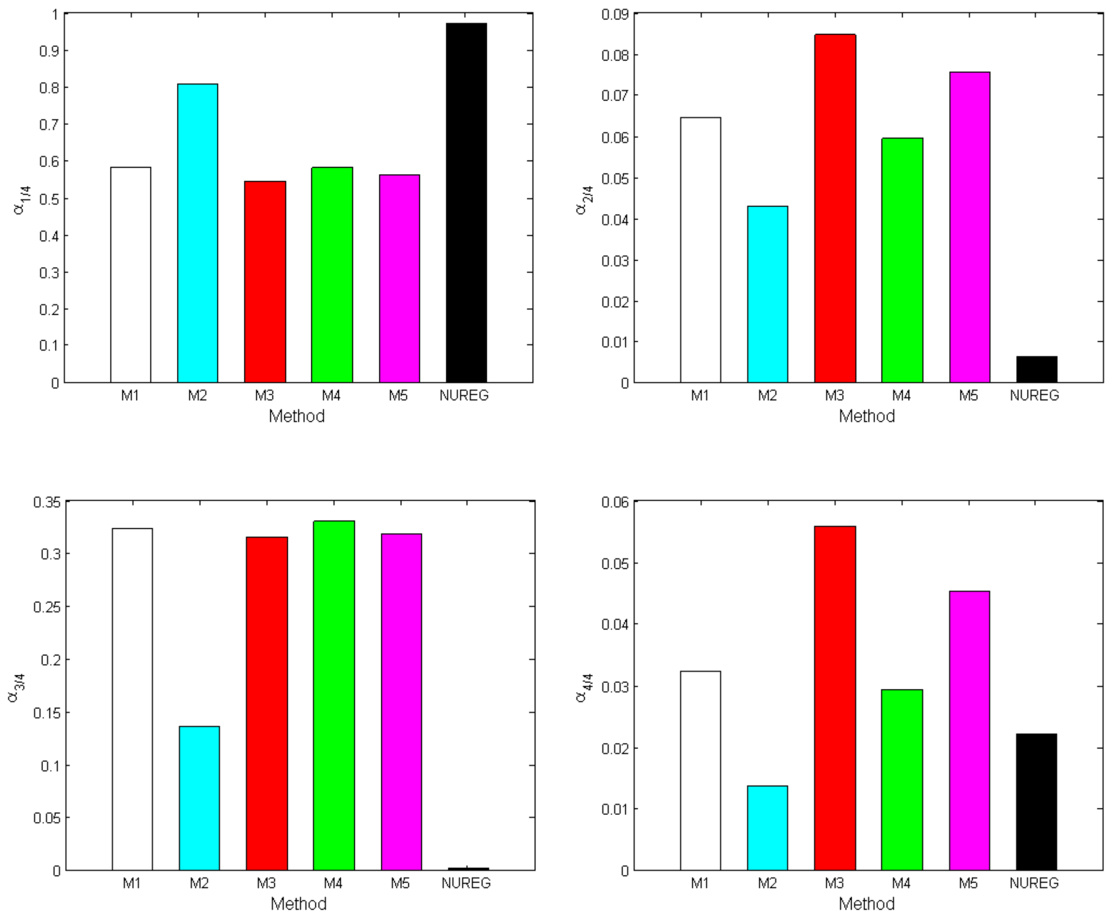


Figure 5.6: Comparison of alpha factors by different methods

- $\alpha_{1/4}$ obtained by MLE without data mapping (M1) and EB (M3-M5) are in close agreement.
- $\alpha_{2/4}$ obtained by all the methods (M1 – M5) is larger than the NUREG value. EB (M5) gives a higher value than MLE (M2). MLE
- $\alpha_{3/4}$ obtained by EB (M5) is higher than MLE with data mapping (M2). All M1-M5 estimates are higher than NUREG value.
- $\alpha_{4/4}$ obtained by EB (M5) is higher than MLE with data mapping (M2). All EB estimates (M3-M5) estimates are higher than NUREG value.
- A possible reason for CCF rate by M1-M5 being higher than NUREG is that the impact vectors are not considered in the present analysis. Because of which, estimates M1-M5 are more pessimistic (or conservative).

5.3 Example 2: Larger MOV Data

The second data set consists of MOV CCG of 2, 4, 8 12, 16 and 24. The operation time for all the systems is 18 years and only one failure mode FO is considered (see Table 5.13).

Table 5.13: CCF data: case study 2

System size	No. of systems	Total Time	No. of failures					
n	N	N*18*12 (month)	$1/n$	$2/n$	$3/n$	$4/n$	$5/n$	$6/n$
2	412	88992	154	8	0	0	0	0
4	332	71712	209	10	12	11	0	0
8	414	89424	331	7	2	0	0	0
12	3	648	28	0	0	3	0	0
16	8	1728	14	2	1	0	0	1
24	2	432	16	2	0	0	0	0

5.3.1 MLE without Data Mapping

Table 5.14: MLE results without data mapping: case study 2

Multiplicity	k	1	2	3	4
No. of failures	$N_{k/4}$	209	10	12	11
Mean of failure rate (/month)	$\Lambda_{k/4}$	2.91×10^{-3}	1.39×10^{-4}	1.67×10^{-4}	1.53×10^{-4}
Standard deviation (/month)	$\sigma(\Lambda_{k/4})$	2.02×10^{-4}	4.41×10^{-5}	4.83×10^{-5}	4.62×10^{-5}
α factor	$\alpha_{k/4}$	8.64×10^{-1}	4.13×10^{-2}	4.96×10^{-2}	4.55×10^{-2}

The standard MLE method is applied to estimate mean and SD of failure rates.

5.3.2 MLE with Data Mapping

Data from CCG of 8, 12, 16 or 24 was mapped down to CCG of 4 using Equation (2.23) or formulas listed in Table 2.3. Data from CCG 2 was mapped up to 4. Using $\rho = 0.2$. Results of data mapping are given in Table 5.15.

Table 5.15: CCF data mapped to CCG of 4: case study 2

System size		Number of failures					
n		$1/n$	$2/n$	$3/n$	$4/n$	$5/n$	$6/n$
2	original	154	8	0	0	0	0
	mapped	154	5.1200	2.5600	0.3200	0	0
4	original	209	10	12	11	0	0
	mapped	209	10	12	11	0	0
8	original	331	7	2	0	0	0
	mapped	170.3571	2.3571	0.1429	0.0000		
12	original	28	0	0	3	0	0
	mapped	10.6909	1.0182	0.1939	0.0061		
16	original	14	2	1	0	0	1
	mapped	5.1670	0.5995	0.1170	0.0082	0	0
24	original	16	2	0	0	0	0
	mapped	3.2464	0.0435	0.0000	0.0000	0	0
sum	mapped	552.4615	19.1383	15.0138	11.3343	For HPP	

Table 5.16: MLE results after data mapping: case study 2

Multiplicity	k	1	2	3	4
Number of failures	$N_{k/4}$	552.4615	19.1383	15.0138	11.3343
Mean of failure rate (/month)	$\Lambda_{k/4}$	2.18×10^{-3}	7.57×10^{-5}	5.94×10^{-5}	4.48×10^{-5}
Standard deviation (/month)	$\sigma(\Lambda_{k/4})$	9.29×10^{-5}	1.73×10^{-5}	1.53×10^{-5}	1.33×10^{-5}
α factor	$\alpha_{k/4}$	9.24×10^{-1}	3.20×10^{-2}	2.51×10^{-2}	1.90×10^{-2}

5.3.3 EB without Data Mapping

Given the data for the target system (CCG=4) in Table 5.13, EB approach was used to obtain results in Table 5.17. The parameters of the prior are $\alpha = 0.3737$ and $\beta = 441.7277$. The prior and posterior distributions of the failure rate are plotted in Figure 5.7.

Table 5.17: Posterior Mean and SD of failure rate without data mapping

Multiplicity	k	1	2	3	4
No. of failures	$N_{k/4}$	209	10	12	11
Mean of failure rate (/month)	$\Lambda_{k/4}$	2.90×10^{-3}	1.44×10^{-4}	1.71×10^{-4}	1.58×10^{-4}
Standard deviation (/month)	$\sigma(\Lambda_{k/4})$	2.01×10^{-4}	4.46×10^{-5}	4.88×10^{-5}	4.67×10^{-5}
α factor	$\alpha_{k/4}$	8.60×10^{-1}	4.26×10^{-2}	5.08×10^{-2}	4.67×10^{-2}

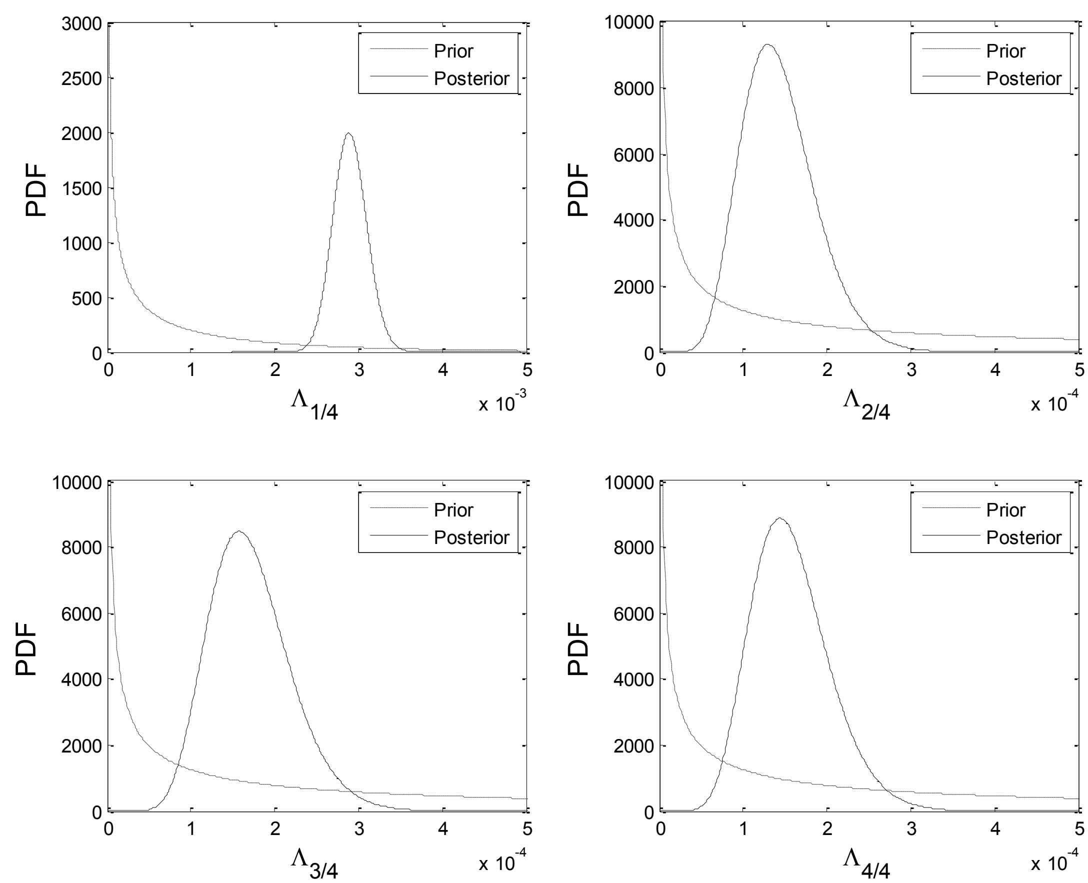


Figure 5.7: Failure rate distributions without data mapping: case study 2

5.3.4 EB with Data Mapping (Multiple Priors)

In order to estimate alpha factors of the 4-MOV system accurately, precise evaluations of the failure rates $\Lambda_{k/4}$ for all multiplicities k are necessary. Consider 1/4 failures of all the four sizes listed in Table 5.15 as an example. The number of failures after data mapping and the corresponding time are listed in the second and third columns of the table below, which are required by the empirical Bayes algorithm. The detailed calculation procedure has been listed Table 4.1. The outcome is the posterior mean value of the failure rate $\Lambda_{1/4}$ of the CCG=4 after Bayesian update which has been included in the fourth column.

Table 5.18: Number and exposure times of 1oo4 failures after data mapping

System size n	$N_{1/4}$	T_n (month)
2	154	88992
4	209	71712
8	170.3571	89424
12	10.6909	648
16	5.1670	1728
24	3.2464	432

Repeating the above procedure for all the multiplicities k respectively, the failure rates $\Lambda_{k/n}$ of the target system (CCG=4) are believed to be closer to the inherent unknown rates. The calculation results are shown in Table 5.19 below.

Table 5.19: Posterior mean and SD of failure rate with data mapping

Multiplicity	k	1	2	3	4
Mean of failure rate (/month)	$\Lambda_{k/4}$	2.92×10^{-3}	1.41×10^{-4}	1.58×10^{-4}	1.44×10^{-4}
Standard deviation (/month)	$\sigma(\Lambda_{k/4})$	2.02×10^{-4}	4.40×10^{-5}	4.45×10^{-5}	4.28×10^{-5}
α factor	$\alpha_{k/4}$	8.68×10^{-1}	4.20×10^{-2}	4.70×10^{-2}	4.29×10^{-2}

Table 5.20: Parameters of priors (mapped data): case study 2

Failure rate	$\Lambda_{1/4}$	$\Lambda_{2/4}$	$\Lambda_{3/4}$	$\Lambda_{4/4}$
α	0.8586	0.2748	0.5790	0.3422
β	167.3990	1098.9120	7914.9329	6893.9587

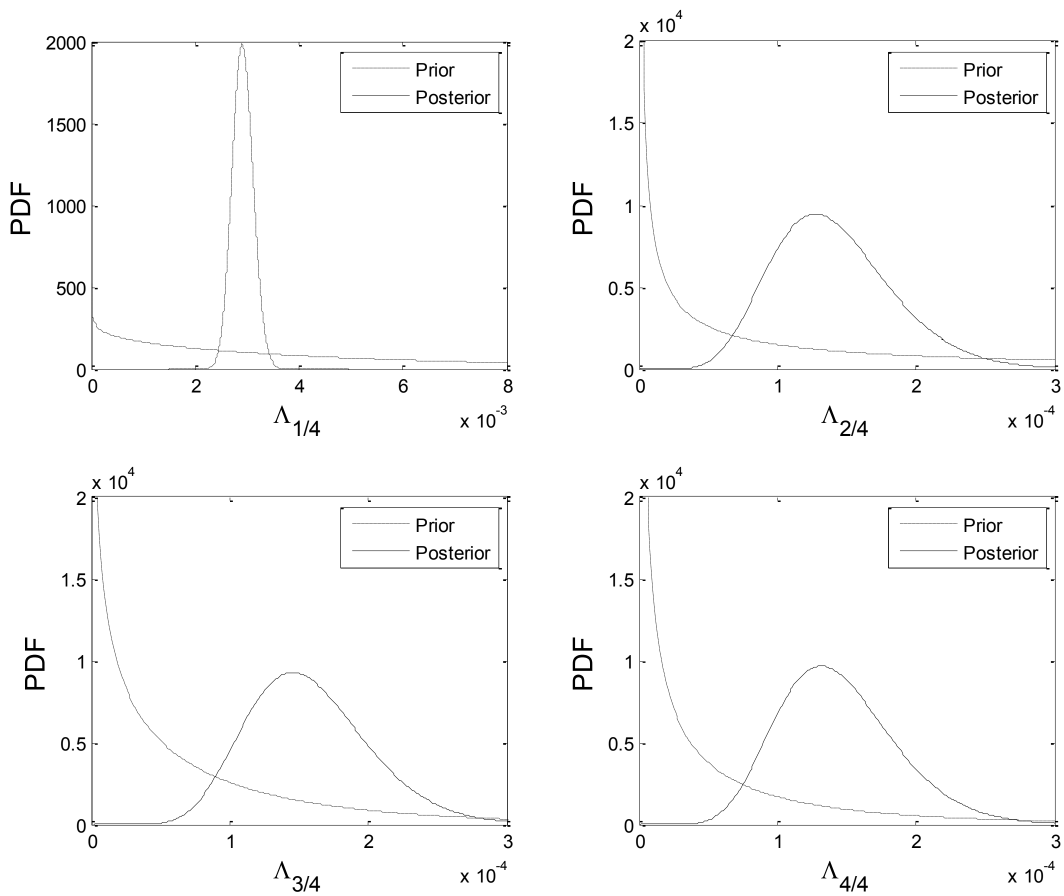


Figure 5.8: Failure rate distributions with data mapping (multiple priors)

5.3.5 EB with Data Mapping (Single Prior)

In this case, a sin gle prior is assigned to data given in Table 5.15. The parameters of the prior are estimated as $\alpha = 0.1712$ and $\beta = 119.6847$, and calculation results are presented in Table 5.21. The prior and posterior distributions of the failure rate are plotted in Figure 5.9.

Table 5.21: Posterior mean and SD of failure rates (mapped data, single prior)

Multiplicity	k	1	2	3	4
No. of failures	$N_{k/4}$	209	10	12	11
Mean of failure rate (/month)	$\Lambda_{k/4}$	2.91×10^{-3}	1.42×10^{-4}	1.69×10^{-4}	1.56×10^{-4}
Standard deviation (/month)	$\sigma(\Lambda_{k/4})$	2.01×10^{-3}	4.44×10^{-5}	4.86×10^{-5}	4.65×10^{-5}
α factor	$\alpha_{k/4}$	8.62×10^{-1}	4.19×10^{-2}	5.02×10^{-2}	4.60×10^{-2}

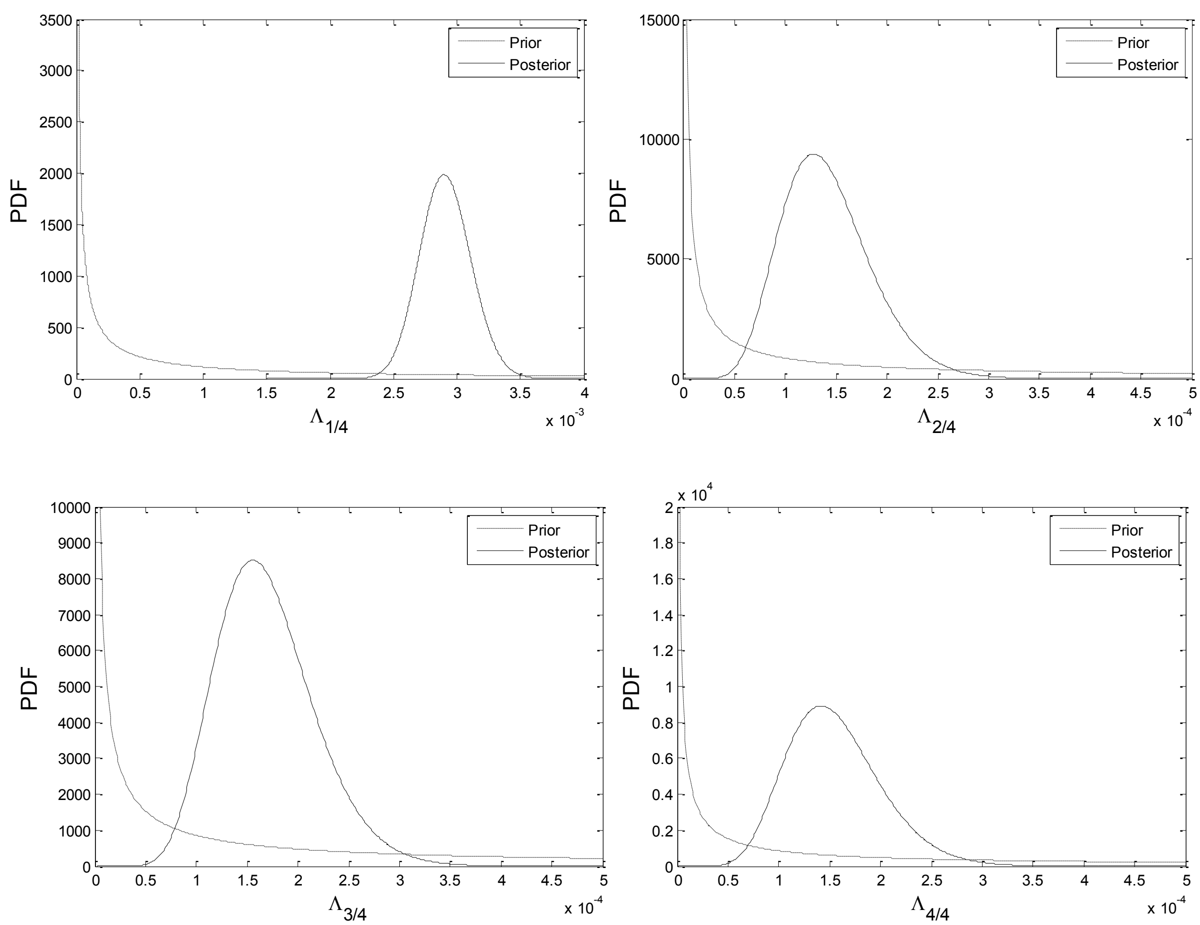


Figure 5.9: Failure rate distributions with data mapping (single prior)

5.3.6 Comparison of Results

Table 5.22 presents a comparison of the estimated mean of failure rates, and a graphical comparison is shown in Figure 5.10.

Table 5.22: Mean failure rates (/month) by different methods (CCG=4)

No.	Method	$\Lambda_{1/4}$	$\Lambda_{2/4}$	$\Lambda_{3/4}$	$\Lambda_{4/4}$
M1	MLE without data mapping	2.91×10^{-3}	1.39×10^{-4}	1.67×10^{-4}	1.53×10^{-4}
M2	MLE with data mapping	2.18×10^{-3}	7.57×10^{-5}	5.94×10^{-5}	4.48×10^{-5}
M3	EB without data mapping	2.90×10^{-3}	1.44×10^{-4}	1.71×10^{-4}	1.58×10^{-4}
M4	EB with data mapping (multiple priors)	2.92×10^{-3}	1.41×10^{-4}	1.58×10^{-4}	1.44×10^{-4}
M5	EB with data mapping (single prior)	2.91×10^{-3}	1.42×10^{-4}	1.69×10^{-4}	1.56×10^{-4}

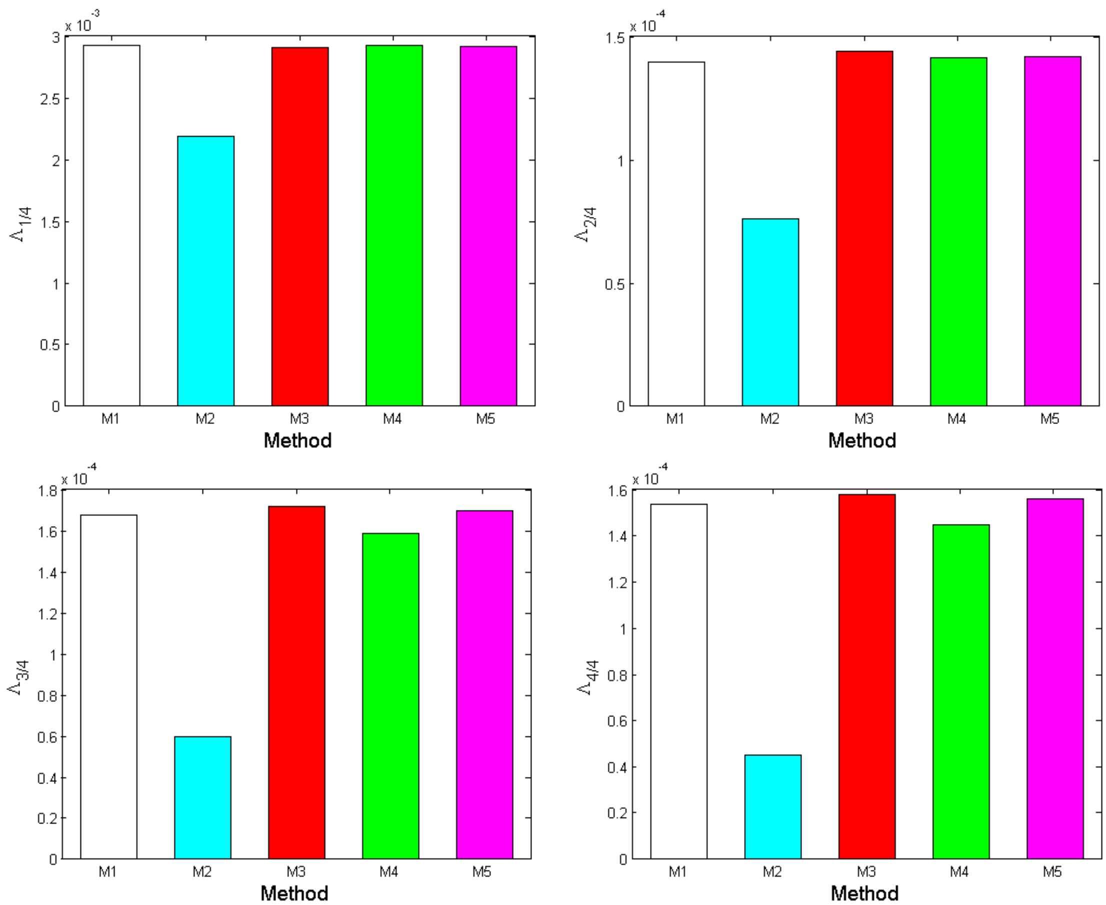


Figure 5.10: Comparison of estimates of mean failure rates (case study 2)

- EB estimates of the mean are quite close to MLE without data mapping (M1).

- MLE with mapped data (M2) leads to lowest value of mean rate in each case.

Table 5.23:SD of failure rates (/month) by different methods: case study 2)

No.	Method	$SD(\Lambda_{1/4})$	$SD(\Lambda_{2/4})$	$SD(\Lambda_{3/4})$	$SD(\Lambda_{4/4})$
M1	MLE without data mapping	2.02×10^{-4}	4.41×10^{-5}	4.83×10^{-5}	4.62×10^{-5}
M2	MLE with data mapping	9.29×10^{-5}	1.73×10^{-5}	1.53×10^{-5}	1.33×10^{-5}
M3	EB without data mapping	2.01×10^{-4}	4.46×10^{-5}	4.88×10^{-5}	4.67×10^{-5}
M4	EB with data mapping (multiple priors)	2.02×10^{-4}	4.40×10^{-5}	4.45×10^{-5}	4.28×10^{-5}
M5	EB with data mapping (single prior)	2.01×10^{-3}	4.44×10^{-5}	4.86×10^{-5}	4.65×10^{-5}

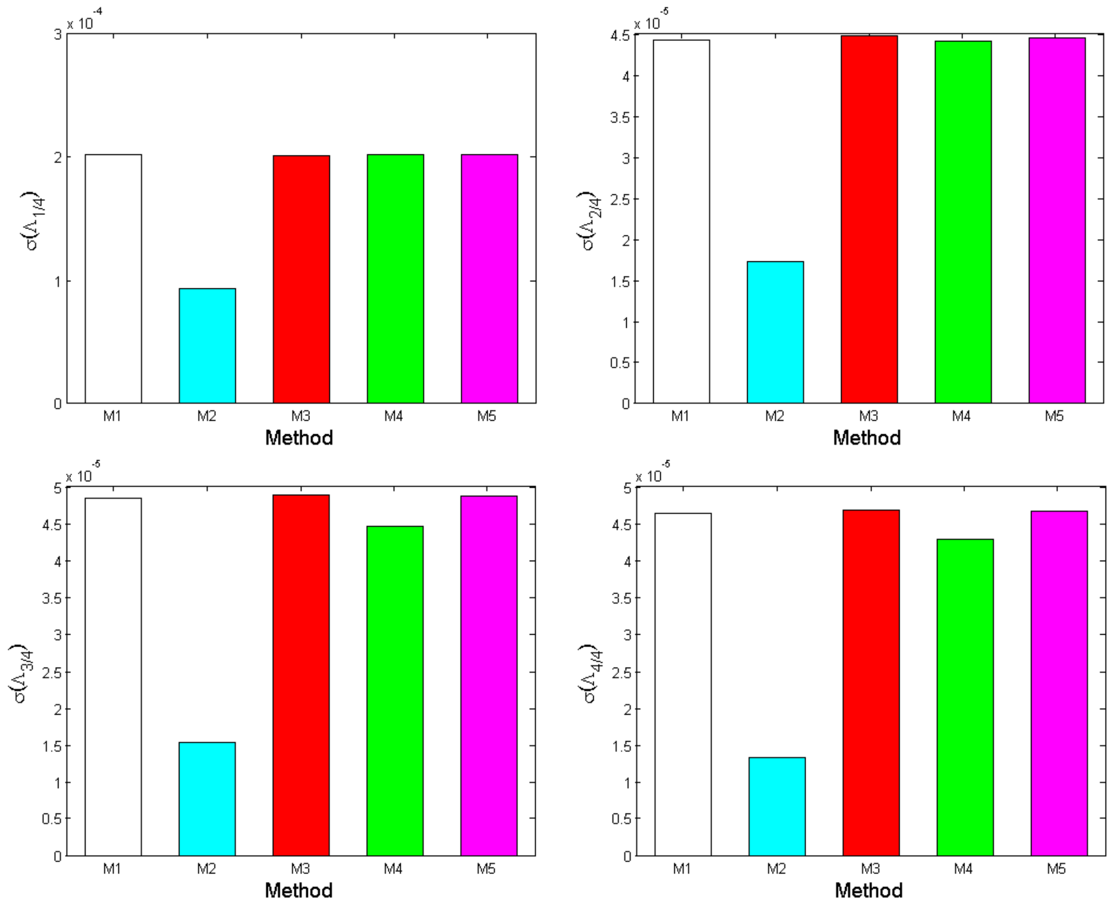


Figure 5.11: Comparison of estimates of SD of failure rates: case study 2

- EB estimates of the SD are quite close to MLE without data mapping (M1).

Table 5.24: Alpha factors by different methods: case study 2

No.	Method	$\alpha_{1/4}$	$\alpha_{2/4}$	$\alpha_{3/4}$	$\alpha_{4/4}$
M1	MLE without data mapping	8.64×10^{-1}	4.13×10^{-2}	4.96×10^{-2}	4.55×10^{-2}
M2	MLE with data mapping	9.24×10^{-1}	3.20×10^{-2}	2.51×10^{-2}	1.90×10^{-2}
M3	EB without data mapping	8.60×10^{-1}	4.26×10^{-2}	5.08×10^{-2}	4.67×10^{-2}
M4	EB with data mapping (multiple priors)	8.68×10^{-1}	4.20×10^{-2}	4.70×10^{-2}	4.29×10^{-2}
M5	EB with data mapping (single prior)	8.62×10^{-1}	4.19×10^{-2}	5.02×10^{-2}	4.60×10^{-2}
Empirical	NUREG/CR-5497	9.69×10^{-1}	6.50×10^{-3}	2.50×10^{-3}	2.22×10^{-2}

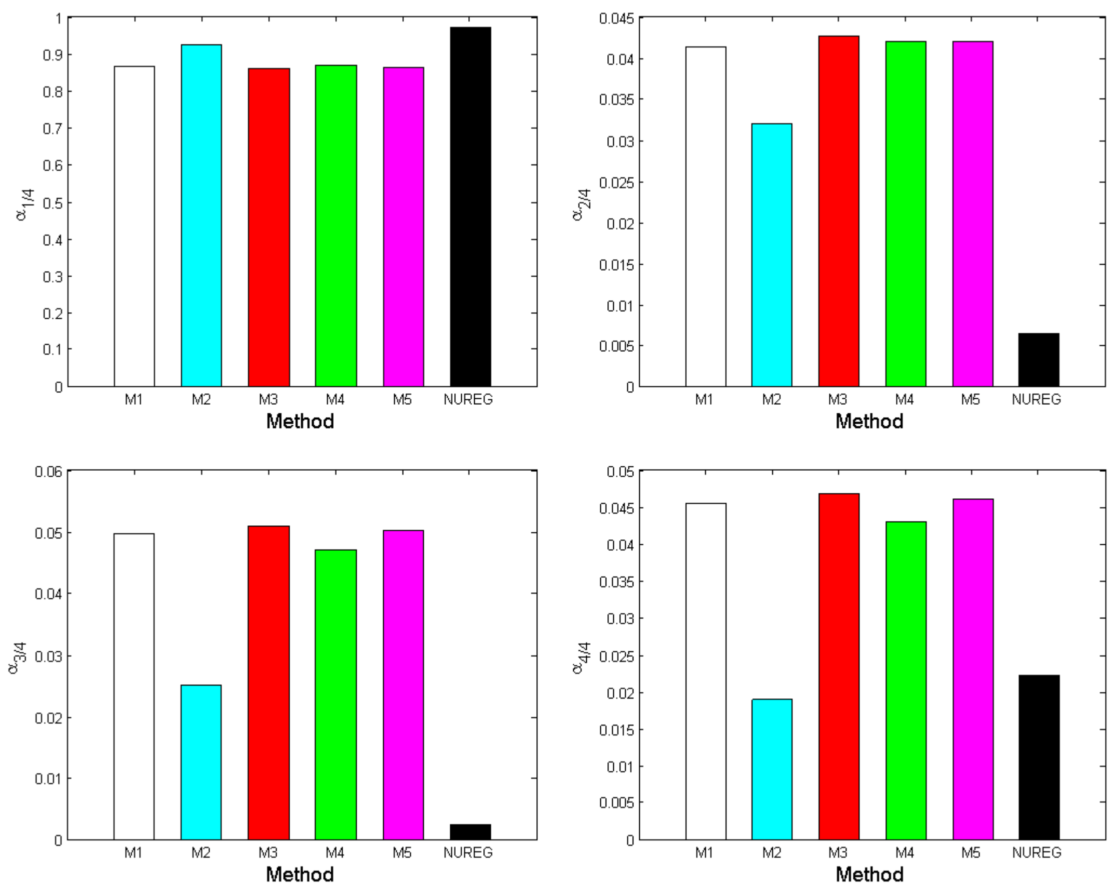


Figure 5.12: Comparison of alpha factors: case study 2

- $\alpha_{1/4}$ obtained by all the methods are in relatively close agreement.
- $\alpha_{2/4}$ and $\alpha_{3/4}$ obtained by all the methods (M1 – M5) is fairly large than the corresponding NUREG values.
- $\alpha_{2/4}$, $\alpha_{3/4}$ and $\alpha_{4/4}$ obtained by EB methods (M3-M5) and MLE without mapping (M1) are in relatively close agreement.

5.4 Summary

This section presents detailed examples of evaluating CCF rates for data related to MOVs. The CCF data can be utilized in 5 different ways depending on whether or not data mapping is done and how Bayesian priors are selected.

The CCF rates estimated in this report are higher (more conservative) than NUREG published values. The reason is that impact vector information has not been included in the analysis, since it was not available. The construction of impact vectors is generally based on expert assessment, which is not available to us.

It is interesting to note that when a data set has a relatively high number of failures (as in case study 2), EB estimates of CCF rates are quite close to that obtained by simple MLE method without data mapping. Thus, EB's utility is apparent only in cases of fairly sparse failure data.

This Section also demonstrates that a capacity has been developed to analyze CCF rates using the Empirical Bayesian (EB), which has been adopted by the Nordic countries of ICDE project.

Section 6:
Summary and Conclusions

This project presents a comprehensive review of the literature related to probabilistic modeling of CCF events and statistical estimation of the model parameters using the EB method adopted by many members of the ICDE project.

6.1 Summary

This research project has achieved the following tasks:

- A comprehensive review of major approaches for modeling the common cause failure data is presented.
- An insightful analysis of the data mapping methods is presented.
- A clear explanation of "General multiple failure rate model" is presented, which is the most current and logically sound method.
- Statistical estimation of the CCF rates and alpha-factors is described.
- Empirical Bayes (EB) method is described and programmed for data analysis.
- Two detailed case studies using MOV data are presented, which serve as tutorial examples for the users.

6.2 Conclusions

- This project demonstrates that a capacity has been developed in Canada to analyze CCF rates using the Empirical Bayesian (EB), which is a method of choice of the Nordic countries participating in the ICDE project.
- The EB method is a conceptually coherent method for pooling the data across the systems within a plant and across the plants. The EB method is explored in detail by considering different combinations of priors.

- Cases studies show that depending on the quantity of data EB results can be close to or far apart from the maximum likelihood estimates.

6.3 Recommendations

- The use of impact vectors should be considered in the statistical analysis.
- The impact of testing scheme, such as staggered testing, on the estimation of CCF rates should be evaluated by proper probabilistic modeling techniques.
- Analysis of CCF data from other safety systems should be undertaken as the data become available.
- The Canadian industry currently uses the Alpha factor published in the US NRC handbooks. The US NRC uses an expert assessment method for the construction of impact vectors. The Canadian industry should be encouraged to utilize EB method and compare the results with those obtained from existing method.

Appendix A
Review of CCF models

This Appendix provides a brief review of CCF models developed in the PSA literature.

1. Beta factor (BF) Model

A variety of parametric CCF models have been developed in the past decades (Fleming *et al*. 1986) such as the Beta factor model, multiple Greek letter model, and Alpha factor model. This Section provides a comprehensive review of these methods.

It is one of the earliest CCF models that originated from a report written by Fleming (1974). The BF model assumes that all the n components fail when a CCF event occurs in a group. Thus, there are only two types of failures: failure of either one component (probability or rate λ_1) or all n failure (probability or rate λ_n). The key part in the model is to introduce a fraction of failure β_s and then define the total failure probability/rate of a certain component λ_{total} as

$$\beta = \frac{\lambda_n}{\lambda_n + \lambda_1} = \frac{N_n}{N_n + (N_1/n)} \quad (1)$$

$$\lambda_{total} = \lambda_n + \lambda_1 = \frac{N_n + (N_1/n)}{N} = \lambda_n/\beta \quad (2)$$

The data required in the BF model are:

N_1 = Number of independent failure events in the group.

N_n = Number of failure events that simultaneously involve n components in the group.

N = Number of total demands applied to the group.

In case of estimation of failure rates, instead of probability of failure, the number N should be replaced by the total system operation time T.

The simplifying assumption of the BF model makes it easy to represent the failure consequences and to compute with the sole factor β defined. However, when the group size n

is larger than 2, the assumption is obviously against the practical situations because the CCF is not necessarily involving all the components within the group.

2. Multiple Greek Letter (MGL) Model

The MGL model is extension of the BF model (Fleming and Kalinowski, 1983) in which a set of different parameters are defined according to the ratio of multi-component CCFs as follows.

λ_{total} = Total failure rate for each component.

β = Conditional probability that a component's failure is shared by one or more additional components, given that the former component fails.

γ = Conditional probability that a component's failure is shared by two or more additional components, given that the former component fails together with one or more additional components.

Estimation formulas for these parameters are given as follows:

$$\lambda_{total} = \frac{1}{nT}\sum_{i=1}^{n} iN_i \tag{3}$$

$$\beta = \sum_{i=2}^{n} iN_i \,/ \sum_{i=1}^{n} iN_i \tag{4}$$

$$\gamma = \sum_{i=3}^{n} iN_i \,/ \sum_{i=2}^{n} iN_i \tag{5}$$

A drawback is that the data may not be sufficient to ensure the accuracy of many factors especially when the group size is large. Moreover, a problem of over-parameterization occurs as the group size increases to a high level.

3. Basic Parameter (BP) Model

The BP model is either time-based (to calculate failure rates) or demand-based (to calculate probability of failure) according to Kumamoto and Henley (1996). The parameter $\lambda_{k/3}$ is failure rate for a particular group of $i(1 \leq k \leq n)$ components. Considering a three-component system for example, the rate of events involving exactly two components is thus $3\lambda_{2/3}$. From symmetry assumption that the probabilities of failures involving the same number of components are equal (e.g. $\lambda_{1,1} = \lambda_{1,2} = \lambda_{1,3}$), the independent and common cause failure rates depend only on the numbers of failures of different types. The required data are listed below and the formula of failure rate is given in the equation below (Fleming *et al*., 1986; Mosleh *et al*., 1989).

$$\lambda_{k/n} = \frac{N_{k/n}}{\binom{n}{k}T} \tag{6}$$

T = Total operation time of the system.

$N_{k/n}$ = Number of failure events involving exactly k components in failed states.

The BP model is a straightforward approach to adopt simple concepts. Since the components in various trains are tested differently and the testing scheme affects the CCF probabilities, it is necessary to investigate the formulas of CCF factors under different testing schemes. Hwang and Kang (2011) developed the calculation equations of CCF factors within a CCCG under the staggered testing scheme, non-staggered testing scheme and mixed testing scheme. It is a significant generalization of the BP model that makes the method more applicable in the CCF modeling.

4. Binomial Failure Rate (BFR) Model

BFR model is specialized from of a general model proposed by Marshall and Olkin (1967). Failures are divided into two types, namely independent failures and nonlethal shocks. The occurrence of the nonlethal shock follows the Poisson process with a rate μ. Each component has a constant failure probability of p under the shocks and the failure distribution is binomial (Vesely, 1977). Afterwards, Atwood developed the BFR model further and introduced lethal shock with rate ω. The basic parameters are shown below:

λ_I = Independent failure rate for each component.

μ = Occurrence rate for nonlethal shocks.

p = Probability of failure for each component under nonlethal shocks. $(q = 1 - p)$

ω = Occurrence rate of lethal shocks.

The relations hold between the BP and BFR models as follows:

$$\lambda_1 = \lambda_I + \mu p q^{n-1} \tag{7}$$

$$\lambda_k = \mu p^k q^{n-k} \tag{8}$$

$$\lambda_n = \mu p^n + \omega \tag{9}$$

An important feature of this model is that the total number of parameters remains constant regardless of the number of components. Each event is classified as lethal or nonlethal (including independent cause) shock (Kumamoto and Henley, 1996). However, μ cannot be estimated directly from the data recorded because nonlethal shocks do not necessarily cause visible failures. Only shocks that cause at least one component failure are counted. Hence, the rate of visible nonlethal shocks, μ_+, is regarded as a basic parameter instead of μ. Then the expected total number of failures caused by nonlethal shocks is

$$N_{non} = \mu T n p = \frac{\mu_+ T n p}{1 - q^n} = \frac{N_+ n p}{1 - q^n} \tag{10}$$

Then the probability p can be calculated by solving Equation (10). The other parameters are easy to estimate following the equations below

$$\lambda_I = \frac{N_I}{nT} \tag{11}$$

$$\mu_+ = \frac{N_+}{T} \tag{12}$$

$$\omega = \frac{N_L}{T} \tag{13}$$

where

N_I = Number of single-component failures due to independent cause.

N_+ = Number of nonlethal shocks causing at least one component failure.

N_L = Number of lethal shocks.

Ina practical situation no one can predict that how components within a CCCG are subjected to a nonlethal shock due to different location, internal structure and aging. Therefore the probability p may not be a constant for all the components.

5. Example: Comparison of CCF Models

Data regarding diesel generator (CCCG = 4) failures is summarized in Table 2.1 (Becker *et al*., 2009). Observation time is 2.91E6 (the unit is unknown in the report). CCF models discussed previously will be applied to compare their performance. Since BF model does not consider k out n failures, it is not included in the case study. First, α factors are calculated following Equation (2.3) to (2.5). With the α factors and total failure rate, it is not hard to obtain the system and component level failure rates.

Table 5.25: Failure data of diesel generators in SKI Report 2009:07.

k	1	2	3	4
$N_{k/4}$	3	19	2	8
$\alpha_{k/4}$	0.09375	0.59375	0.0625	0.25

Second, the MGL, AF, BP, and BFR models are compared for estimating the failure rate of the system given *k-out-of*-4 criteria (as defined in Section 1). By following the formulas of each model, it is not difficult to obtain the component level and system level failure rates. Simplified expressions for the *k-out-of*-4 system failure rates are able to obtain as well. The estimation of the system failure rates are listed below.

Table 5.26: Failure rates of a k-out-of-4 system calculated by different models.

Model	$k=1$	$k=2$	$k=3$	$k=4$

MGL	1.10E-05	9.96E-06	3.44E-06	2.75E-06
Alpha	1.10E-05	9.96E-06	3.44E-06	2.75E-06
BP	1.10E-05	9.96E-06	3.44E-06	2.75E-06
BFR	1.09E-05	8.94E-06	6.61E-06	3.91E-06

An interesting phenomenon shows up that the MGL, Alpha and BP models produce the same estimates. This result stems from the fact that parameters of MGL and Alpha models are defined by those of BP model. BFR model generates a higher estimate for the 4-*out-of*-4 system. The reason is that non-lethal shock also contributes to the lethal failure with a probability of p.

6. Other Approaches

6.1 Markov Model

All the models above are developed for non-repairable systems. As to the repairable systems, Markov transition diagrams are used with appropriate parameters: common cause occurrence rates $\lambda_1 \sim \lambda_n$ and repair rate μ (Platz 1984).

Let P_i, a function of time, denote the probability of the state i. The probability can transit from state i to state j. A transfer matrix A consisting of $\lambda_1 \sim \lambda_n$ and μ can be formed, the element of which isa_{ij}.The probability of transiting to state j during the intervaldt is $a_{ij}dt$. Then differential equations for P_i can be written in the form of matrix. Then the probability P_i can be calculated by solving the differential equations.

$$P_i(t + \Delta t) = AP_i(t)dt \tag{14}$$

This report does not concentrate on solving differential equations as the Markov model has not been considered suitable for practical applications.

6.2 Unified Partial Method (UPM)

Since the parametric models involve many simplifying assumption and the lack of CCF data, the UK nuclear industry are currently adopting the unified partial method (UPM). In the

UPM common cause method (Zitrou and Bedford, 2003), the defenses against CCF are broken down into 8 sub-factors such as Operator Interaction, Redundancy, Analysis, etc. Each sub-factor has 5 levels of strength (A, B, C, D, E) accompanied by corresponding scores from low to high.

From the database, a beta factor for a specific system i can be calculated using the following equation,

$$\beta_i = \frac{N_{i\_e}}{N_{i\_c}} \tag{15}$$

where N_{i_e} and N_{i_c} are the numbers of common cause events and failed components respectively. Next, a linear regression model is assumed to investigate the relationships between the beta factor and the scores (x_{ij}) corresponding to the eight sub-factor levels.

$$\beta_i = w_1 x_{i1} + w_2 x_{i2} + \cdots + w_8 x_{i8} \tag{16}$$

After the weights (w_i) have been determined, the dependencies of failure rate on the sub-factors are also revealed. And the failure rate of the system λ_i is assumed to be the sum of the eight scores x_i.

The advantage of the UPM is its ease of implementation framework. The steps are clear to follow and the calculation is not difficult; a number of tables are enough to obtain the scores of the sub-factors for each system. However, the data required such as N_{i_e} and N_{i_c}, and x_i are not available at most times because not many countries are utilizing this method currently. The failure events and data may not be recorded in such a detail as the UPM expects and the scores have to be determined by the experts following some guides and past experience. Besides, any change of the design and age of the system are likely to require rescoring the system. A lot of qualitative analysis is required when analysts are assigning the scores. Moreover, the assumed linear relationship between failure rate and the scores needs further investigation. The levels of certain sub-factors may change the other factors' impacts on CCF probability according to Zitrou and Bedford (2003).

6.3 Influence Diagram (ID) Extension

The Bayesian network (BN) model is a directed acyclic graph in which the nodes represent random variables and the directed arrows always point from parents to children. All the conditional probability functions (CPFs) for the variables are written in such a conditional way: $f(x_i|pa(x_i))$ where $pa(x_i)$means the parents of a variable x_i have already happened. It is necessary to choose an appropriate family of distribution for each CPF and assign values to the parameters. In this process, empirical data and/or expert judgments may be utilized (Langseth and Portinale, 2007).

An obvious advantage of the BN model lies in the flexibility of the formulation. The parent-child conditional structure enables analysts to add possible causes of failures into the model. What's more, since the number of probabilities is reduced significantly, the data or expert judgments required are reduced which are usually expensive and difficult to obtain. The difficulty of the method is that the analyst should get inputs from the real experts because of the complexity of the networks. All the possible dependencies of the variables should be considered in order to reflect the practical situations. This makes it rather difficult to construct a BN structure.

As an extension of the Bayesian network, influence diagram (ID) can be used to modify the UPM (Zitrou, Bedford and Walls, 2004). The ID contains different types of nodes (variables) such as decision node, uncertainty node and value node, which is more detailed compared to the BN. It has two major advantages in the CCF defense analysis. First, the ID graphically portrays the dependencies among the defenses and the influences on the CCF probability. Second, the expert judgments can be taken into account together with the statistical data.

The assumption made in the model is that the CCF event occurs in an independent homogeneous Poisson process (HPP) with rate γ_i and hence the overall failure rate λ is the superposition of them. It is also assumed that the parameters of rates are uncertain variables rather than unknown constants.

There are two factors causing CCFs, (i.e., root cause and coupling mechanisms, three defense actions are proposed aimed at reducing the frequency of root causes, or against the coupling mechanisms, or both. And the domain variables are classified to three types related to different objectives, namely defenses, root causes and coupling factors. Three scenarios are illustrated: a), decision variable describes the defense against the occurrence of root cause; b), decision variable describes the defense against the coupling factor; and c), decision variable describes the defense against both of them.

Zitrou *et al*. (2007) further introduce the ID extension of the UPM afterwards. Two advantages are shown in the application of the ID, namely modeling non-linear dependencies amongst the defenses and taking account for expert judgments as an important source. First, functional interactions between any two defenses can be modeled by assuming parameters to represent impacts on a certain variable between different levels of the defenses. The relationships may be functionally independent, functionally dependent and threshold functionally dependent. Second, when constructing the ID network, every expert is expected to draw the relationships between defense, root cause and coupling mechanism variables individually. Then all the opinions are incorporated and disagreements can be discussed within the panel.

One of the most important advantages of the ID is the capability of representing various kinds of information (e.g., root cause, coupling factor, defense, rate, etc.) by different types of variables. Besides, the dependencies between various defenses can be taken into account. Moreover, expert judgments are valued as an important resource. However, different interpretations, personal experience and ambiguities in definitions of the defenses may lead to different understandings of the interacti0ons. This will lead to several rounds of reflections so that the experts in the panel can reach an agreement in the end.

Appendix B
Derivation of Vaurio's EB estimator

Vaurio (1987) proposes a procedure to estimate the hyper-parameters using a moment matching method. More specifically, since the mean and variance of the prior are able to be expressed as α/β and α/β^2, if one can calculate them accurately, then α and β are obtained as well.

A set of normalized weight w_i are introduced into the procedure which sum up to one. A weighted average of the sample mean M_0 is calculated by the following equation:

$$M_0 = \sum_{i=1}^{n} w_i N_i / T_i \tag{1}$$

which is an unbiased estimate of the true mean of prior distribution, denoted as M.

The variance of the sample mean is calculated as

$$V_0 = S + M_0 / T^* \tag{2}$$

where

$$S = 1/(1 - \sum_{i=1}^{n} {w_i}^2) \sum_{i=1}^{n} w_i (N_i/T_i - M_0)^2 \tag{3}$$

$$T^* = \sum_{i=1}^{n} T_i - max\ (T_i) \tag{4}$$

in order to avoid the problem caused by identical data that the component with largest observation time has the highest precision while the others do not.

Based on the method of moments, the author concludes that V_0 is a biased estimate of the true variance V and derives the variance of the estimated sample mean as follows.

$$Var(M_0) = V\sum_{i=1}^{n} w_i{}^2 + M\sum_{i=1}^{n}(w_i{}^2/T_i) \tag{5}$$

The normalized weights are generated by minimizing the variance of sample mean as shown above. The formula is

$$w_i = u_i/\sum_{j=1}^{n} u_j \tag{6}$$

where

$$u_i = T_i/(T_i + M_0/V_0) \tag{7}$$

However, since the true values of M and V are never known in practice, one can use the estimates of them, M_0 and V_0, and iterate from Equation (1) to (7) with a set of initial weights as $w_i = 1/n$. When M_0, V_0 and w_i converge, the iteration can be terminated and the values of M_0 and V_0 can be seen as the real ones of M and V. With the converged prior mean and variance, calculation procedures of α and β have been given as follows according to Vaurio (1987).

$$\beta = M/V \tag{8}$$

$$\alpha = M^2/V + 0.5\beta/T^* \tag{9}$$

There is a similarity between the EB and JS estimators. First, transform the posterior mean of the EB estimator into the form of stein estimator.

$$\lambda_i = \frac{\alpha + N_i}{\beta + T_i} = \frac{\alpha}{\beta}\frac{\beta}{\beta + T_i} + \frac{N_i}{T_i}\frac{T_i}{\beta + T_i} \tag{10}$$

According to the prior distribution, the mean and variance are expressed as $M = \alpha/\beta$ and $V = \alpha/\beta^2$. Then the hyper-parameter can be obtained in the form of the statistical characters as $\beta = M/V$. Substitute β into the equation above, one can come up with a new expression without hyper-parameters.

$$\lambda_i = M\frac{M/T_i}{M/T_i + V} + \frac{N_i}{T_i}\frac{V}{M/T_i + V} \tag{11}$$

Based on the derivations given by Vaurio (1987), M/T_i can be seen as the variance of likelihood distribution while V is the variance of prior distribution. Therefore, the EB estimator mean value can be expressed in the form of a balance between the overall average and MLE with weights determined by the magnitudes of prior and likelihood's variances.

$$\lambda_i = M\frac{Var(Likelihood)}{Var(Likelihood) + Var(Prior)} + \frac{N_i}{T_i}\frac{Var(Prior)}{Var(Likelihood) + Var(Prior)} \tag{12}$$

Second, recall the formulas of the JS estimator.

$$\lambda_i = B_i\mu + (1 - B_i)N_i/T_i$$

where $$B_i = \frac{\mu}{\mu+\sigma^2 T_i} \tag{13}$$

The estimator is also able to be written as a balance between overall average and MLE as shown in Equation (13) below. The fractions in this formula are also proportions taken by likelihood and prior in the total variance based on the derivations in the literature (Vaurio and Jänkälä, 1992) and they are exactly the same with those in the EB mean value expression.

$$\lambda_i = \mu\frac{\mu}{\mu + \sigma^2 T_i} + \frac{N_i}{T_i}\frac{\sigma^2 T_i}{\mu + \sigma^2 T_i} = \mu\frac{\mu/\;_{\mathrm{i}}}{\mu/\;_{\mathrm{i}} + \sigma^2} + \frac{N_i}{T_i}\frac{\sigma^2}{\mu/\;_{\mathrm{i}} + \sigma^2} \tag{13}$$

As a conclusion, there is a close connection between the EB and JS estimators, although the JS estimator does not need to specify the exact form of prior distribution of the failure rates.

References

Arnold, S. (1990) *Mathematical statistics*. New Jersey: Prentice-Hall.

Atwood, C. L. (1986) The binomial failure rate common cause model. *Technometrics* 28, pp. 139-148.

Becker, G., Jänkälä, K., Johanson, G., *etc*. (2007) SKI Report 2007:41, *Dependency Analysis Guidance, Nordic/German Working group on Common Cause Failure analysis* - Phase 1 project report: Comparisons and application to test cases.

Becker, G., Johanson, G., Lindberg, S. and Vaurio J. K. (2009) SKI Report 2009:07, *Dependency Analysis Guidance, Nordic/German Working group on Common Cause Failure analysis* - Phase 2, Development of harmonized approach and applications for common cause failure quantification.

Börcsök, J. and Holub, P. (2008) Different approaches for probability of common cause failure on demand calculations for safety integrity systems. *IEEE/ACS International Conference on Computer Systems and Applications* (AICCSA), Doha, Quatar.

Carlin, B. P. and Louis T. A. (2000) *Bayes and Empirical Bayes Methods for Data Analysis*.2nd edition. London: Chapman & Hall/CRC.

Crowder, M. J., Kimber, A. C., Smith, R. L. and Sweeting, T. J. (1991) *Statistical Analysis of Reliability Data*. Chapman & Hall, London.

Efron, B. (2010) Chapter 1: Empirical Bayes and James-Stein estimator. *Large-scale simultaneous inference*. http://www-stat.stanford.edu/~omkar/329/chapter1.pdf

Efron, B. and Morris, C. (1975) Data analysis using Stein's estimator and its generalizations. *Journal of the American Statistical Association*, Vol 70, No. 350, pp. 311-319.

Evans, M. G. K., Parry, G. W. and Wreathall, J. (1983) On the treatment of common-cause failures in system analysis. *Reliability Engineering* 9, pp. 107-115.

Fleming, K. N. (1974) A reliability model for common mode failure in redundant safety systems. *General Atomic Report*, *GA-13284*, Pittsburgh, PA.

Fleming, K. N., Mosleh, A. and Deremer, R. K. (1986) A systematic procedure for the incorporation of common cause event, into risk and reliability models. *Nuclear Engineering and Design* 58, pp. 415-424.

Fleming, K.N. and Kalinowski, A.M. (1983) An Extension of the Beta Factor Method to Systems with High Levels of Redundancy. *Report PLG-0289*, Pickard, Lowe and Garrick.

Hanley, J. A. and Lippman-Hand, A. (1983) If nothing goes wrong, is everything all right? Journal of the American Medical Association, Vol 249, No. 13, pp. 1743–1745.

Hokstad, P. (2004) A generalization of the beta factor model. *In Proceedings ESREL 2004/PSAM 7*,pp. 1363-1368.

Hokstad, P. and Rausand, M. (2008) Common cause failure modeling: status and trends. *In Handbook of PerformabilityEngineering*, Chapter 39, pp.621–640.

Hwang, M. J. and Kang, D. I. (2011) Estimation of the common cause failure probabilities on the component group with mixed testing scheme. *Nuclear Engineering and Design*241, Issue 12, pp. 5275 – 5280.

James, W. and Stein, C. (1961) Estimation with quadratic loss. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*. Vol. 1, Berkeley: University of California Press, pp. 361-379.

Karlis, D. (2002) http://www.stat-athens.aueb.gr/~karlis/multivariate%20Poisson%20models.pdf

Kumamoto, H. and Henley, E. J. (1996) *Probabilistic Risk Assessment and Management for Engineers and Scientists*. New York: IEEE Press.

Kvam, P. H. and Miller, J. G. (2002) Common cause failure prediction using data mapping. *Reliability Engineering and System Safety* 76, pp. 273-278.

Langseth, H. and Portinale, L. (2007) Bayesian networks in reliability. *Reliability Engineering and System Safety* 92, pp. 92-108.

Mao S. S. (1999) *Bayesian Statistics*. ISBN 7-5037-2930-9.

Marshall, A. W. and Olkin, I. (1967) A multivariate exponential distribution. *Journal of the American Statistical Association* 62, pp. 30-44.

Marshall, F. M., Rasmuson, D. M. and Mosleh, A. (1998) *Common-Cause Failure Parameter Estimations*. NUREG/CR-5497.Washington, DC: U.S. Nuclear Regulatory Commission.

Mosleh, A. and Siu, N.O. (1987) A multi-parameter event-based common cause failure model. *Trans. 9th International Conference Structural Mechanics in Reactor Technology*, Lausanne, Switzerland, Vol. M.

Mosleh, A., Fleming, K. N., Parry, G. W., Paula, H. M., Worledge, D. H., Rasmuson, D. M. (1989) *Procedures for Treating Common Cause Failures in Safety and Reliability Studies*. NUREG/CR-4780.Washington, DC: U.S. Nuclear Regulatory Commission.

Pandey, M. D. and Jyrkama, M. I. (2012) *CIVE-601 Engineering Risk and Reliability*.

Platz, O. (1984) A Markov model for common-cause failures. *Reliability Engineering* 9, pp. 25-31.

Quigley, J., Bedford, T. and Walls, L. (2007) Estimating rate of occurrence of rare events with empirical Bayes: A railway application. *Reliability Engineering and System Safety* 92, pp. 619-627.

Shen, Y. (2006) Master Thesis: *Discussion and application of the empirical Bayes method in multi-parameter estimation problem*. http://www.doc88.com/p-878818503572.html

Stott, J. E., Britton, B. T., Ring, R. W., Hark, F. and Hatfield, G.S. (2010) Common cause failure modeling: aerospace vs. nuclear. *Proceedings of the 10th International Conference on Probabilistic Safety Assessment and Management*, Seattle, Washington, USA.

Technical Brief from Relex Software Corporation. Calculating MTTF when you have zero failures.http://microblog.routed.net/wpcontent/uploads/2006/08/calculating_mttf_with_zero_failures.pdf

Vaurio, J. K. (1987) On analytic empirical Bayes estimation of failure rates. *Risk Analysis*. Vol 17, No. 3, pp. 329-338.

Vaurio, J. K. (1995) The Probabilistic modeling of external common cause failure shocks in redundant systems. *Reliability Engineering and System Safety* 50, pp. 97-107.

Vaurio, J. K. (2007) Consistent mapping of common cause failure rates and alpha factors. *Reliability Engineering and System Safety* 92, pp. 628-645.

Vaurio, J. K. (2008) Common Cause Failure Modeling. *Encyclopedia of Quantitative Risk Analysis and Assessment*. http://onlinelibrary.wiley.com/doi/10.1002/9780470061596.risk0497/pdf

Vaurio, J. K. and Jänkälä, K.E. (1992) James-Stein estimators for failure rates and probabilities. *Reliability Engineering and System Safety* 36,pp. 35-39.

Vaurio, J. K. and Lindén, G. (1986) On robust methods for failure rate estimation. *Reliability* Engineering 14, pp. 123-132.

Vesely, W. E. (1977) Estimating common cause failure probabilities in reliability and risk analyses: Marshall-Olkinspecializations. *In Proc. of the Int. Conf. on Nuclear Systems Reliability Engineering and Risk Assessment*, pp.314-341.

Voet, H. V. D. (1994) Comparing the predictive accuracy of models using a simple randomization test. *Chemometrics and Intelligent Laboratory Systems* 25, pp. 313-323.

Xing, L. M., Fleming, K. N.andLoh, W. T. (1996) Comparison of Markov model and fault tree approach in determining initiating event frequency for systems with two train configurations. *Reliability Engineering and System Safety* 53, pp. 17-29.

Zitrou, A, Bedford, T. and Walls, L. (2004) Developing soft factors inputs to common cause failure Models. *Probabilistic safety assessment and management*.Springer, Berlin. pp.825-830.

Zitrou, A. and Bedford, T. (2003) Foundations of the UPM common cause method. *Safety and Reliability*. Balkema, ESREL 2003.pp. 1769-1775.

Zitrou, A., Bedford, T. and Walls L. (2007) An influence diagram extension of the unified partial method for common cause failures. *Quality Technology and Quantitative Management*. Vol. 4, No. 1, pp. 111-128.