Statistical Analysis of Common Cause Failure Events using ICDE Data

S. Yu[a], M.D. Pandey[a], S. Yalaoui[b] and Y. Akl[b]
[a] University of Waterloo, Waterloo, Canada
[b] Canadian Nuclear Safety Commission, Ottawa, Canada

Abstract: 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 paper utilizes a modern method, called "General Multiple Failure Rate Model", for the probabilistic modeling of CCF events. To estimate the parameters of the GMFR model, the Empirical Bayes (EB) method is adopted. A detailed case study is presented using CCF data for Motor Operated Valves (MOVs).

Keywords:
PSA, Common cause failure, Poisson Process, Alpha factor, ICDE Data base, Motor Operated valves

1. INTRODUCTION

In the 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 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 [1, 2]. This paper describes the General Failure Rate model of CCF events, and presents a detailed case study to evaluate the model parameters using the Empirical Bayes (EB) method along with the data mapping techniques. The case study is based on CCF data for motor operated valves (MOVs).

2. BASIC MODELING OF CCF EVENTS

The probabilistic basis for CCF modelling is that the occurrences of failures in a single component follows 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

$$P[N(t) = k] = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \tag{1}$$

where λ denotes the failure rate, defined as the average number of failures per unit time.
Historically, several conceptual probabilistic models of CCF events have been presented in the literature. The General Multiple Failure Rate (GMFR) model has been a new and widely accepted model of CCF events, which is also advocated by the working group of the Nordic countries [3].
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 = $\sum_{k=1}^{n} \Lambda_{k/n}$

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

$$\Lambda_{k/n} = \binom{n}{k} \lambda_{k/n} = \frac{N_{k/n}}{T} \quad (2)$$

$$\Lambda_n = \sum_{k=1}^{n} \binom{n}{k} \lambda_{k/n} = \sum_{k=1}^{n} \Lambda_{k/n} \quad (3)$$

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

$$\alpha_{k/n} = \frac{\Lambda_{k/n}}{\Lambda_n} \quad (4)$$

3. ESTIMATION OF CCF RATES

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 [4]:

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

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, $\pi(\lambda)$, is assigned to the failure rate λ, which is usually estimated from past experience and expert judgment. The conjugate prior, the gamma distribution, is a common choice due to mathematical simplicity and its flexibility to fit various types of data [5].

$$\pi(\lambda_i) = \frac{\beta^{\alpha} {\lambda_i}^{\alpha-1} e^{-\beta\lambda_i}}{\Gamma(\alpha)} \quad , \qquad \lambda_i, \alpha, \beta > 0 \quad (6)$$

The mean and variance of the prior are

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

Poisson likelihood for the failure events is given as

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

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

$$p(\lambda_i|N_i, T_i) = \frac{(\beta + T_i)^{(\alpha+N_i)} {\lambda_i}^{\alpha+N_i-1} e^{-(\beta+T_i)\lambda_i}}{\Gamma(\alpha + N_i)} \quad (7)$$

The mean and variance of the gamma posterior are given as

$$\widehat{M}_i = (\alpha + N_i)/(\beta + T_i) \text{ and } \hat{V}_i = (\alpha + N_i)/(\beta + T_i)^2 \quad (8)$$

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

Empirical Bayes (EB) is a method for estimating parameters of a prior distribution used in the Bayesian analysis. In this paper, the EB method proposed by Vaurio [6] has been adopted, since it is also followed 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.

In PSA, methods have been developed to assimilate CCF data available from systems of various sizes to analyse 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. The mapping down means the mapping of the source data from systems with CCG greater than n to a target k-oo-n system. In this paper, a mapping down method proposed by [7] has been adopted. The mapping up means the mapping of the source data from systems with CCG less than n to a target k-oo-n system. In this paper, a mapping up method proposed by Mosleh *et al*. [8] has been adopted.

2. CASE STUDY: MOV DATA

This Section presents a comprehensive case study 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 data set consists of CCF data from MOV CCG of 2, 4, 8 and 16, as shown in Table 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.
The objective of the case study is to estimates CCF rates using this data for a target system consisting of a parallel system of 4 MOVs. The statistical analysis is based on MLE and EB methods.

Table 1: CCF Data for MOVs

System size	No. of systems	Total Time	No. of failures					
n	m	$m \times 18 \times 12$ (months)	$1/n$	$2/n$	$3/n$	$4/n$	$5/n$	$6/n$
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

2.1. MLE without Data Mapping

In this case only the data for CCG=4 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 Λ_(k/4), and corresponding α-factors are given in Table 2.

Table 2: MLE results for CCF rates

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}

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 3.

Table 2: 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	

In MLE analysis, all the mapped number of failures of a particular multiplicity *k-oo-n* 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 4. The comparison of results obtained with and without mapping is given in the next Section of the paper.

Table 3: MLE results with data mapping

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}

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.

Table 4: 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}

2.4. EB with Data Mapping: Multiple Priors

In this case mapped data given in Table 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: 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 3 is analyzed. The number of failures after data mapping and the corresponding exposure time are given in Table 6. The EB analysis leads to the posterior mean of $\Lambda_{1/4}$ as 4.49×10^{-3} failures per month.

Table 6: 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 6. Parameters of the gamma prior are given in Table 8.

Table 7: 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

2.5. Empirical Bayes (EB) with Data Mapping: Single Priors

In this case, it is assumed that a single gamma prior is applicable to entire mapped data given in Table 2, 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 9 and distributions are plotted in Figure 1.

Table 8: 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}

3. COMPARISON OF RESULTS OF THE CASE STUDY

3.1. Comparison of mean failure rate

In order to understand the effect assumptions associated with 5 estimation methods, the mean of the failure rates are compared in Table 10 and graphically shown in Figure 2.

Table 9: Mean failure rates (per month) by different methods used in the case study

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 1: Failure rate distributions with data mapping (EB with single prior)

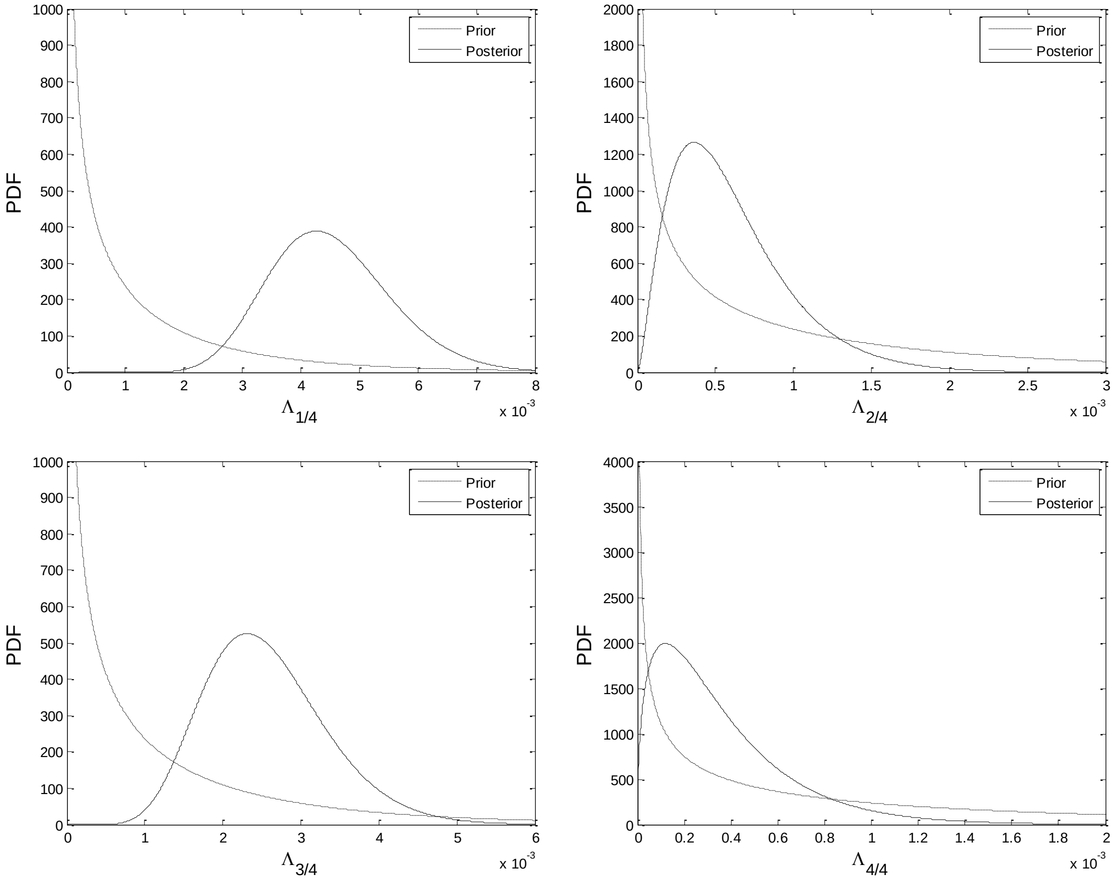
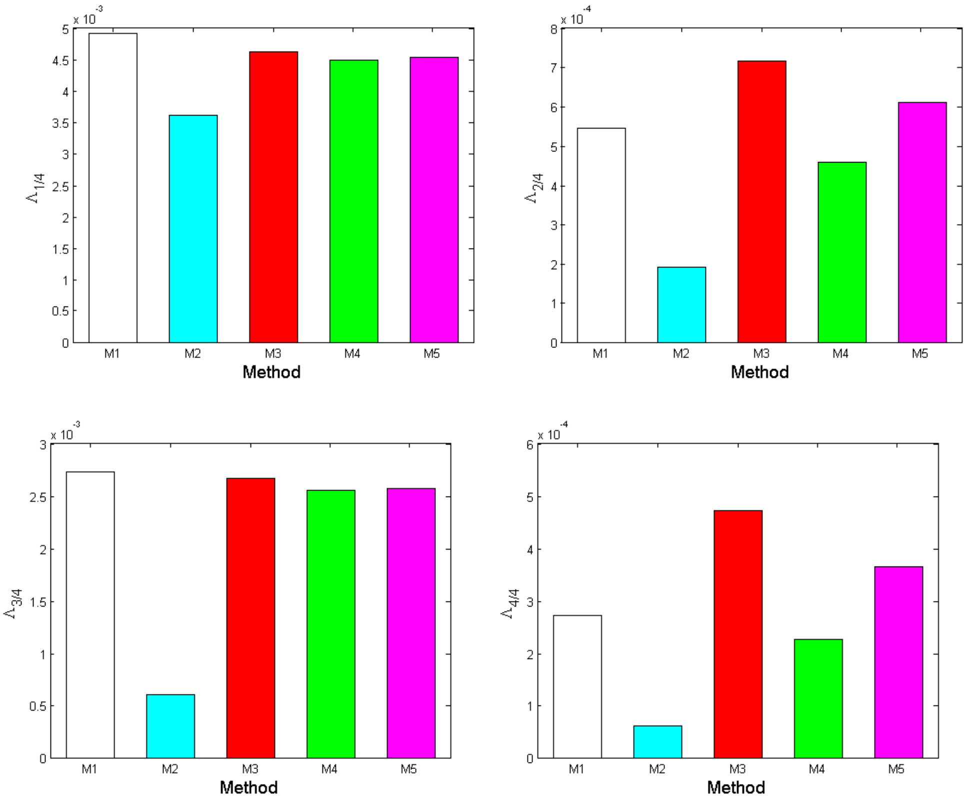


Figure 2: Comparison of mean failure rates estimated in the case study



The following observations are notable:

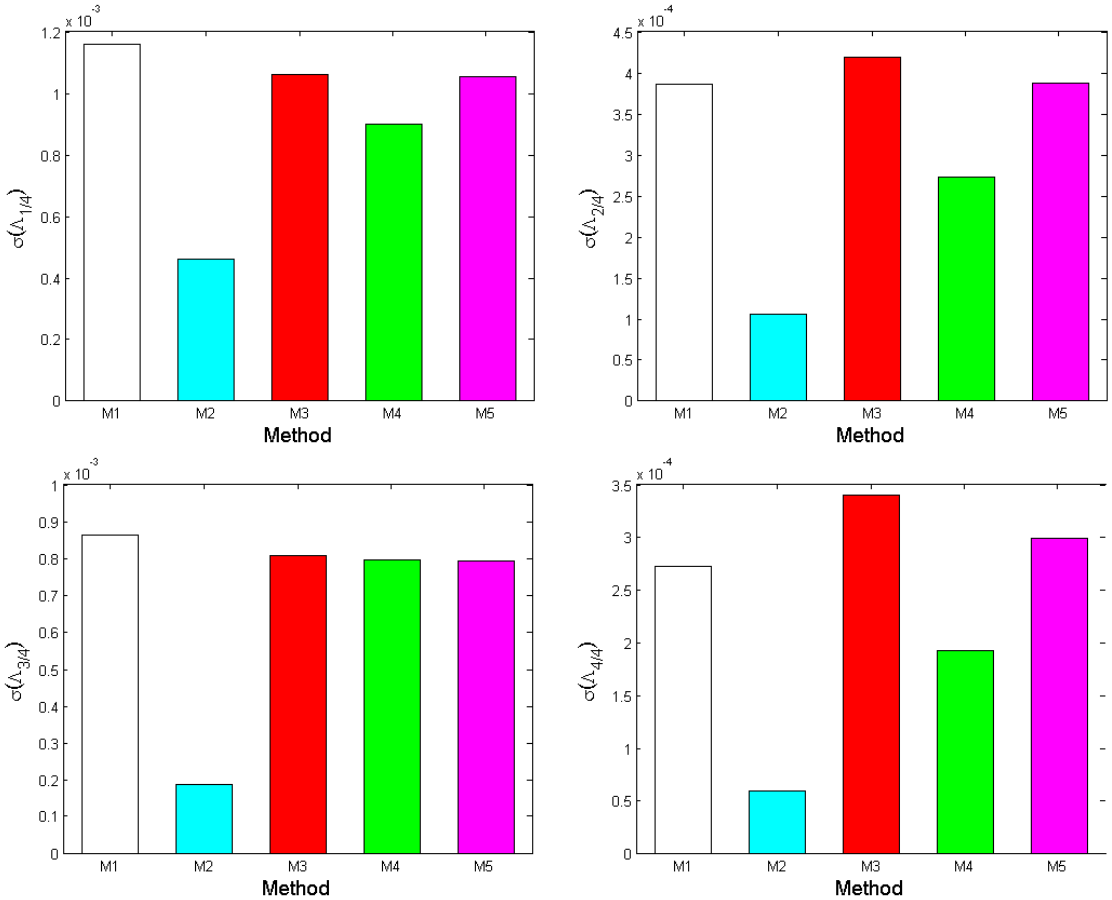
- MLE with data mapping (M2) leads to lower values of mean failure rate as compared to that without data mapping (M1).
- EB with data mapping and single prior (M5) leads to higher mean rate than MLE with data mapping (m2).

3.2. Comparison of standard deviation of failure rate

Table 10: Standard deviation of failure rates by different methods

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 3: Comparison of standard deviations of failure rates by different methods



Key observations are

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

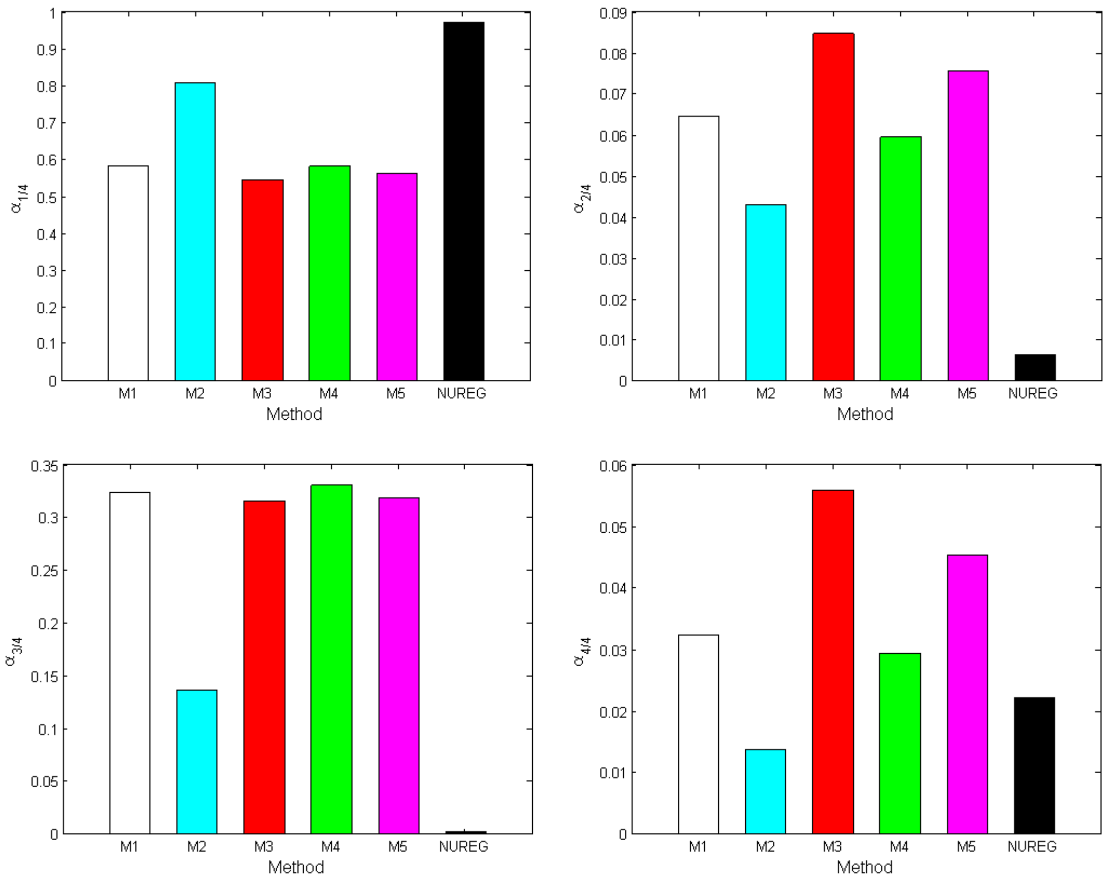
3.3. Comparison of Alpha factors

In Table 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 4.

Table 11: Alpha factors by different method

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}

Figure 4: Comparison of alpha factors by different methods



Key observations are as follows:

- $\alpha_{1/4}$ 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 more conservative).

4. CONCLUSION

This paper describes the basic background of probabilistic modelling of CCF events and the estimation of parameters using MLE and EB methods. A case study is presented in which CCF data related to MOVs are analyzed in detail. The CCF data can be utilized in 5 different ways depending on whether or not data mapping is done and how the Bayesian priors are selected.

It is interesting to note that when a data set has a relatively high number of failure events, EB estimates of CCF rates turn out to be quite close to those obtained by simple MLE method without data mapping. Thus, EB's utility is apparent only in cases of fairly sparse failure data.

References

[1] 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.

[2] 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.

[3] 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.

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

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

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

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

[8] 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.