Developing Generic Prior Distributions for Common Cause Failure Alpha Factors

Zhegang Ma*, John Schroeder, and Curtis Smith
Idaho National Laboratory, Idaho Falls, USA

Abstract: Common cause failures (CCFs) have been recognized as significant risk contributors since the early launching of probabilistic risk assessments (PRAs). A series of United States Nuclear Regulatory Commission (U.S. NRC) regulation (NUREG) reports have been published since the 1980s to provide guidelines for CCF modeling in PRA. A CCF database system has been developed and maintained by the NRC and Idaho National Laboratory (INL) for the nuclear industry. However, although the CCF database system has been routinely maintained and the CCF parameter estimations updated on a yearly basis, the process for developing CCF parameter prior distributions has not been published and the prior distributions themselves have not been updated since the early 2000s. In this paper, an overview of the history of the NUREG CCF reports is provided. Existing CCF prior distributions are listed. The process for developing prior distributions for CCF parameter estimations is described. Prior distributions for CCF alpha factor are updated with data from 1997 through 2015. The issues identified from the study as well as the suggested future work are discussed.

Keywords: PRA, Common Cause Failure, Parameter Estimation, Prior Distribution, Alpha Factors.

1. INTRODUCTION

Common cause failures (CCFs) have been recognized as significant risk contributors since the early launching of probabilistic risk assessments (PRAs) for commercial nuclear power plants. Since 1980s, a series of reports including those of United States (U.S.) Nuclear Regulatory Commission (NRC) regulations (NUREGs) have been published to provide guidelines for performing CCF modeling using probabilistic risk assessment and performing CCF event data analysis. A CCF database system has been developed and maintained by the NRC and Idaho National Laboratory (INL) for the U.S. commercial nuclear power industry. The CCF database system includes a CCF database that stores coded CCF events and a CCF software that uses impact vectors and mapping methods to estimate CCF parameters from the events stored in the CCF database. Generic prior distributions were developed and included in the CCF software for CCF Alpha Factor Model (AFM) parameter estimations. However, while the CCF database has been maintained since its development in late 1990s and the CCF parameter estimations have been updated and published on a yearly basis, the process for developing prior distributions has not been published and the prior distributions themselves have not been updated since the early 2000s. This paper reviews the history of the NUREG CCF related reports (Section 2), presents the existing CCF alpha factor prior distributions (Section 3), and describes the current process to develop prior distributions (Section 4). Section 5 updates the prior distributions for CCF alpha factors using the current process with data from 1997 through 2015. Section 6 discusses the issues encountered during the prior distribution development. Section 7 presents conclusions and suggests the work to do in the future to improve the CCF prior development process.

2. AN OVERVIEW OF NUREG CCF REPORTS

This section provides an overview of the key NUREG reports on the development of CCF modeling guidelines and the NRC CCF database system.

* Zhegang.Ma@inl.gov.

NUREG/CR-4780, also EPRI NP-5613, *Procedures for Treating Common Cause Failures in Safety and Reliability Studies*, Volumes 1 and 2 [1, 2], was published in January 1988 to present the framework for including CCFs in risk and reliability evaluations. It provides procedures to perform and document CCF analysis with a practical and systematic approach. The framework includes the following four major stages: (1) system logic model development, (2) identification of common cause component groups, (3) common cause modeling and data analysis, and (4) system quantification and interpretation of results. While it is not the purpose of the report "to advance or promote a particular method or technique," it does introduce the concept of impact vectors for CCF event classification and representation, and a mapping method that adjusts the original impact vectors to account for common cause group size differences in common cause parameter estimation. Appendix D of the report provides a detailed discussion of the background and justification for using the mapping method in parameter estimation. Although there have been concerns about the mapping method especially for the mapping up technique (when the component group size in the original system is smaller than in the system being analyzed) [3], the use of impact vectors and mapping methods has been adopted in the subsequent NRC CCF studies for CCF parameter estimation.

NUREG/CR-6268, *Common Cause Failure Database and Analysis System*, Volumes 1 to 4 [4, 5], was published in June 1998 to extend previous CCF studies by introducing a method to collect industry failure data, identify and characterize CCF events, and estimate CCF parameters and uncertainties with a computer software. The report uses two data sources at the time for CCF event identification: the Nuclear Plant Reliability Data System (NPRDS), which contains component failure information, and the Sequence Coding and Search System (SCSS), which contains Licensee Event Reports (LERs). Data from 1980 through 1995 were analyzed. The report describes the process for the analyst to consistently code CCF events. A CCF database system was developed with a CCF database that can be searched to obtain the CCF events of interest and a CCF software that can be used to estimate CCF parameters. The CCF software stores CCF events, independent failure counts, and estimate CCF parameters for Alpha Factor and Multiple Greek Letter Models based on the CCF event impact vectors and the mapping method.

NUREG/CR-5497, *Common-Cause Failure Parameter Estimations* [6], was published in October 1998 to document the quantitative results of the CCF data collection effort described in Volumes 1 to 4 of NUREG/CR-6268, as well as the insights from the CCF data analysis. It contains the CCF parameter estimates for the majority of the risk important safety systems and components in commercial nuclear power plants.

NUREG/CR-5485, *Guidelines on Modeling Common-Cause Failures in Probabilistic Risk Assessment* [7], was published in November 1998 to provide a set of guidelines to help PRA analysts in modeling CCF events in commercial nuclear power plants. The report brings together the key aspects of the procedural guidelines presented in previous NRC CCF reports, provides additional insights obtained from the CCF applications, and describes the CCF software capabilities and how to apply the CCF database information to PRA studies.

NUREG/CR-6268, Revision 1, *Common Cause Failure Database and Analysis System, Event Data Collection, Classification, and Coding* [8], was published in September 2007 to update the guidance for collecting, classifying, and coding CCF events described in its older version published in 1998. Three data sources are used to select equipment failure reports to be reviewed for CCF event identification: (1) the NPRDS, which contains component failure information from 1980 through 1996; (2) the Equipment Performance and Information Exchange (EPIX), which contains component failure information since 1997; and (3) LER Search, which contains LERs. The updated CCF data analysis includes the following steps: collection of source data, identification of CCF events, coding of CCF events, database quality assurance, data analysis, and parameter estimation. The CCF event information and the independent event count are entered into the CCF database along with the quality assurance verification. The CCF software system uses the impact vector and mapping method to estimate CCF parameters. The impact vector method used in the process is based on physical characteristics of the event that include component degradation, timing, and shared cause. The software allows the analyst to modify generic event impact factors for plant-specific applications, including using the mapping method to account for differences in common cause component group (CCCG) size.

A series of NRC reports were published on the NRC web site, *http://nrcoe.inl.gov/resultsdb*, starting 2003 to update the CCF parameter estimations in NUREG/CR-5497 on a periodic basis. Below is a list of these update reports including the date range of the data used for the update.

CCF Parameter Estimation 2003 Update [9] reflects the version of the CCF database that contains data from 1980 to 2003. However, it uses a starting date of 1/1/1985 in order to avoid using the large number of CCF events in the 1980–1984 period as the trend is decreasing significantly from 1980 to 1985. The analysis also found that the previously recommended value for the mapping up factor, rho, of 0.85 was very conservative. A recommended value of 0.50 for rho was used in the 2003 Update.

CCF Parameter Estimation 2005 Update [10] reflects the version of the CCF database that contains data from 1980 to 2005. It uses a starting date of 1/1/1991 in order to avoid using the large number of CCF events in the 1980–1990 period as the trend is decreasing significantly from 1980 to 1991. The CCF prior distributions developed in this study are used in the current CCF software for estimating CCF parameters.

CCF Parameter Estimation 2007 Update [11] reflects the version of the CCF database that contains data from 1980 to 2007. It still uses a starting date of 1/1/1991 to avoid using the large number of CCF events in the 1980–1990 period.

CCF Parameter Estimation 2009 Update [12] reflects the version of the CCF database that contains data from 1997 to 2009. The starting date is 1/1/1997. The large number of CCF events in the 1980–1996 period are excluded from the analysis (and subsequent analyses) as the trend is decreasing significantly from 1980 to 1997.

CCF Parameter Estimation 2010 Update [13] reflects the version of the CCF database that contains data from 1997 to 2010. The starting date is 1/1/1997.

CCF Parameter Estimation 2012 Update [14] reflects the version of the CCF database that contains data from 1997 to 2012. The starting date is 1/1/1997.

CCF Parameter Estimation 2015 Update [15] reflects the version of the CCF database that contains data from 1997 to 2015. The starting date is 1/1/1997.

It is worthwhile to note that during the development and maintenance of the CCF database system, the whole process of data classification, loading, and parameter estimation have several levels of quality control. For example, all events are reviewed by two INL data analysts to make sure that the events are classified as CCF events and coded correctly. Then a PRA analyst reviews the CCF events and results for consistency and comparison with PRA experience. A final review is performed by independent CCF expert(s) outside of INL. The independent review is usually conducted by CCF experts from an industry organization such as Pressurized Water Reactor Owner Group (PWROG), formerly Westinghouse User Group (WUG).

Nonetheless, the CCF event identification and characterization are still subject to engineering judgement as analysts could have different interpretations of the events and make different assumptions on the mission information from both the event reports and physical and operational descriptions of the plants involved. The uncertainty caused by the data, as well as other uncertainties such as statistical uncertainty and modeling uncertainty, should be identified and properly addressed in CCF study and applications.

3. EXISTING PRIOR DISTRIBUTIONS IN CCF PARAMETER ESTIMATIONS

To develop uncertainty distributions for CCF parameters, a Bayesian estimation procedure is used in which the choice of prior distribution becomes critical. The prior distribution could be the analyst's subjective judgement or based on observed ranges of variation of the parameters. NUREG/CR-5485 [7] discusses the data uncertainty in the development of a statistical database from CCF event reports. Several different approaches for developing prior distributions are mentioned in the report:

(1) Using the hierarchical Bayes method to develop a plant-to-plant variability distribution of various alpha factors (or other CCF model parameters) across all components and failure modes.

(2) Obtaining the maximum likelihood estimate (MLE) for a given alpha factor, and then using a constrained noninformative prior as its uncertainty distribution. This distribution maximizes the uncertainty given a constraint on the mean value and is usually broader than the corresponding hierarchical Bayes distribution.

(3) Using information from the constrained noninformative prior distributions to obtain an estimate of the parameters of a Dirichlet distribution, which can be combined to obtain an effective estimate for the Dirichlet distribution parameters.

(4) Using a mapping method to develop prior distributions for alpha factors for each common cause component group (CCCG) size in an effort to utilize all CCF events in the CCF database. In this approach, all CCF events are mapped to a given CCCG size. The MLE for each alpha factor is obtained and fit with constrained noninformative distribution. The estimates of the Dirichlet distribution parameters are calculated and combined to obtain an effective estimate.

NUREG/CR-5485 uses the last approach, i.e., a mapping method, to develop prior distributions for alpha factors. However, the details of the process have not been documented in the report or published otherwise. Instead, a NRC white paper titled "*Estimation of Industry-Wide Common-Cause Failure Prior Distributions*" with a date of January 2010 might be the best documentation so far that describes the process for estimating CCF prior distributions. The paper used the CCF data from 1995 through 2005 to develop the prior distributions with step-by-step instructions.

Other than the prior distribution results in the 2010 paper and the one in NUREG/CR-5485, there are 3 other formal prior distributions that are documented in the NRC CCF Parameter Estimation Update Reports: 2003 version as in the 2003 Update [9], 2005 version as in the 2005 Update [10], and 2007 version as in the 2007 and subsequent Updates [11 through 15]. Table 1 shows the date range of the data and part of the mean α values in each of the prior distributions as published in these reports.

On the other hand, the prior distributions used for CCF parameter calculations are embedded in the CCF software as a hard-coded table. The table used in the current CCF software was compared with those in the 2003, 2005, and 2007 Updates and found that the software/database uses the same prior distributions as in the 2005 Update, not the latest one as reported in the 2007 Update.

4. PROCESS FOR DEVELOPING GENERIC PRIOR DISTRIBUTIONS

The 2010 White Paper presents the following process to develop CCF prior distribution using an industry-wide data set:

Step 1. For each common cause component group size, tabulate the number of CCF events and complete CCF events (a complete CCF event is one in which all the components in the group fail, with all P_i = 1.0, timing factor =1.0, and shared cause factor =1.0 in the CCF database.

Table 1: Mean Values of Existing CCF Prior Parameters

Parameter	NUREG/CR-5485	2003 CCF Update	2005 CCF Update	2007 CCF Update	2010 White Paper	2009, 2010, 2012, and 2015 CCF Update
α2 (CCCG=2)	4.70E-02	3.09E-02	4.06E-02	2.57E-02	1.75E-02	The prior distributions published in these annual CCF Update reports are the same one as in the 2007 CCF Update
α3 (CCCG=3)	2.58E-02	7.17E-03	8.71E-03	5.79E-03	5.94E-03	
α4 (CCCG=4)	1.86E-02	3.72E-03	4.64E-03	2.98E-03	3.81E-03	
α5 (CCCG=5)	1.46E-02	6.26E-04	7.25E-04	5.33E-04	9.32E-04	
α6 (CCCG=6)	1.23E-02	6.15E-04	6.86E-04	4.07E-04	5.06E-04	
α7 (CCCG=7)	1.03E-02	1.29E-04	1.52E-04	1.17E-04	2.22E-04	
α8 (CCCG=8)	9.06E-03	1.38E-04	1.46E-04	1.25E-04	1.88E-04	
Date Range of Data	1980–1995	1985–2003	1991–2005	1991–2007	1995–2005	1997–2009 (2010,2012,2015)

Step 2. Calculate the n_k's for each group size (2 to 16) using all partial CCF events (a partial CCF event is one in which not all of the components fail). This will involve mapping up and mapping down.

Step 3. Using the information obtained in Step 1, perform a binomial regression to obtain the probability of CCF events in a given group size.

Step 4. Obtain the estimated number of complete CCF events using the results obtained in Step 3. Add this number to the final n_k for each group size. For example for group size 2 add the number to n_2, for group size 4 add the number to n_4.

Step 5. Using the final n_k values, estimate the mean value alpha factors for each group size.

Step 6. Using these final n_i values, estimate the beta prior distributions for each group size. The parameters of the beta distribution are α and β. The beta distribution is denoted by Beta(α, β). A computer code, CalcPrior, was developed by INL to estimate the distributions using a procedure to calculate Dirichlet distribution parameters with noninformative prior distributions.

Step 7. As a check, calculate the mean of each prior distribution and compare with the values obtained in Step 5. The mean value is obtained using the formula $\mu = \alpha / (\alpha+\beta)$.

The main difference between this process and the short descriptions in NUREG/CR-5485 seems to be that the process in the 2010 White Paper separates the complete CCF events from the partial CCF events. While the mapping impact vector method is used for partial CCF events, binomial regression is used to curve fit the complete CCF events. This is probably due to the concern that the mapping method might be adding too many pseudo complete CCF events to other group sizes when the observed complete CCF events in one group size.

5. UPDATING GENERIC PRIOR DISTRIBUTIONS FOR ALPHA FACTORS

This section uses the process described in Section 4 to update the generic prior distributions for alpha factors using data from 1997 through 2015.

5.1. Accessing CCF Data

There are two ways to access CCF data stored in the NRC CCF Database system for analysis. As a general user, one can utilize the CCF Database website https://rads.inl.gov/Pages/CCF.aspx, in which various CCF rules can be defined to choose the CCF date range and other CCF event characteristics such as interested component types, failure modes, and failure causes. As a database administrator, one can query the CCF Database directly with the Structured Query Language (SQL) and output the results to an Excel spreadsheet for further analysis. The first method, or the database website method, can be used to obtain the interested CCF events by selecting CCF event characteristics. The software can generate the number of CCF events and the effective independent event count that satisfy the selection criteria. It can provide the original (or un-mapped) impact vectors for each of the selected CCF events, the mapped impact vectors and adjusted independent count for different group sizes. The impact vector results can be output for further analysis. The second method, or the SQL method, is more powerful as it can dig in and extract any field from the database. Both methods were used to obtain necessary data for this study. Table 2 shows examples of CCF events with their un-mapped impact vectors obtained from the CCF Database website. Table 3 shows the same CCF events that are obtained from the CCF Database SQL querying.

Table 2: Examples of CCF Events Obtained from CCF Database Website

CCFID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
219-1997-0278	0	1	0	0	--	--	--	--	--	--	--	--	--	--	--	--
219-1998-0207	1.8	0.1	--	--	--	--	--	--	--	--	--	--	--	--	--	--
219-1999-0248	0	1	0	0	--	--	--	--	--	--	--	--	--	--	--	--
219-2000-0051	0	1	0	0	0	0	0	0	--	--	--	--	--	--	--	--
219-2003-0369	0	0	1	0	0	0	--	--	--	--	--	--	--	--	--	--
219-2005-0341	0	0	1	0	0	0	--	--	--	--	--	--	--	--	--	--
219-2014-0488	0	1	0	0	0	--	--	--	--	--	--	--	--	--	--	--
220-2001-0398	0	0	0	1	--	--	--	--	--	--	--	--	--	--	--	--
220-2007-0144	0	0	0	1					--	--	--	--	--	--	--	--
220-2010-0412	0	1	0	0	--	--	--	--	--	--	--	--	--	--	--	--
237-1998-0219	1.45	0.025	--	--	--	--	--	--	--	--	--	--	--	--	--	--
237-2004-0336	0	1	0	0	0	0	--	--	--	--	--	--	--	--	--	--

Table 3: Examples of CCF Events Obtained from CCF Database SQL Querying

CCF Name	CCCG	Shock Type	Fail Degree	CCF Cause	System	FM	Comp Type	Descript
219-1997-0278	4	NL	Partial	Design	SWS	FH	MDP	ESW Pump Overhaul Performed With Incorrect Material.
219-1998-0207	2	NL	Almost	Component	EPS	FS	GEN	EDGs Failed Due to Faulty Relays
219-1999-0248	4	NL	Partial	Component	RHR	FS	MDP	Blown fuse affected both pumps on the RHR System II side.
219-2000-0051	8	NL	Partial	Design	ACP	CC	CRB	Aux Xfrmr Feeder Breakers Failed to Trip When Turbine Tripped
219-2003-0369	6	NL	Partial	Environment	CWS	FO	STR	Debris issues at the intake due to high wind operate
219-2005-0341	6	NL	Partial	Environment	CWS	PG	STR	Grass loading of trash racks fail the TSA
219-2014-0488	5	NL	Partial	Component	MSS	AO	SRV	Two of five EMRV actuators failed to operate.
220-2001-0398	4	NL	Complete	Design	ISO	OO	AOV	4 ISO Drain Valve AOVs unable to close.
220-2007-0144	4	NL	Complete	Other	LCS	SA	MOV	Spurious opening of four Core Spray inside isolation valves.
220-2010-0412	4	NL	Partial	Human	MSS	SA	MSV	Two MSIVs spuriously close
237-1998-0219	2	NL	Partial	Human	DCP	FO	BCH	Loose connections on 125 VDC battery chargers.
237-2004-0336	6	NL	Partial	Component	CWS	FO	STR	TSA Screens Not Rotating

CCF data from 1997 through 2015 is chosen for this study as it represents the most recent data and reflects more current plant conditions and practices at the time when the analysis was performed. Also the earliest year that could be chosen from the CCF Database website is 1997. In the CCF Database website, the following selection criteria are defined:

- Type of CCF Event Level: All Level CCF Events
- CCF Event Type: CCF Events Only
- Date Range: 1997 through 2015
- Filter Independent Events by Selected Cause(s): True
- Shock Criteria: All Events
- Redundancy Range: Minimum = 2, Maximum = 16
- Bayesian Update Method: Mean Method
- Failure Modes: select all failure modes except Setpoint
- Plants/Systems/Components/CCF Categories: no selection on these CCF event characteristics

There are a total of 268 CCF events and 7492.8 effective independent failure events from the above selection criteria.

Additional criterion on CCF Categories → Degree → Almost/Partial or Complete are used to obtain the partial CCF events and complete CCF events as required in the existing process. The un-mapped impact vectors and mapped impact vectors are also acquired from the CCF Database website. The mapped impact vectors for partial CCF events for each group size obtained from the website are used directly in the study.

Table 4 shows the number of partial CCF events, number of complete CCF events, and total number of CCF events. Table 5 shows the mapped impact vectors for partial CCF events for each group size obtained from the CCF Database website[†].

5.2. Treating Complete CCF Events

As described in Section 4, the current process treats complete CCF events differently by using the binomial regression method rather than the mapping technique, which is probably due to the concern that the mapping technique might be adding too many pseudo complete CCF events to other group sizes when mapping the observed complete CCF events in one group size. For example, Table 4 shows that there are 34 complete CCF events for group size 2, 12 for group size 3, and 56 for all group sizes. Using the mapping technique, all complete CCF events in group sizes 3 to 16 (which is 56-34=22) would be mapped down to add 22 complete pseudo CCF events to group size 2. For group size 3, all complete CCF events in group sizes 4 to 16 (which is 56-34-12=10) would be mapped down and the complete CCF events in group size 2 (which is 34) would be mapped up. Assuming the conditional probability of failure of each component given a nonlethal shock, ρ, is 0.5, there would have 10+34*0.5=27 pseudo complete CCF events being added to group size 3.

The binomial regression method would not add such number of pseudo complete CCF events as it curve-fits the fraction of complete CCF events over the total number of CCF events. With the binomial regression method [16, 17], *P(m)* is defined as the probability that a CCF event is a complete failure in a group size *m*. It then uses the observed fractions of complete CCF failures in all group sizes and fit the data using a pre-defined function. In this study, MATLAB [18] was used for curve fitting with the following function:

$$\ln\left(\frac{P(m)}{1-P(m)}\right) = a + b\left(1 - e^{-m}\right) \tag{1}$$

[†] While this study is conducted for group sizes from 2 to 16, only the results for group sizes from 2 to 8 are shown in this paper due to size limitation.

Table 4: CCF Data from 1997 through 2015

Group Size	No. Partial CCF Events	No. Complete CCF Events	Total No. CCF Events	Prob. of Complete CCF Event - Data	Prob. of Complete CCF Event - Curve Fitting	Estimated No. Complete CCF Events
2	27	34	61	0.55738	0.51050	31.14031
3	27	12	39	0.30769	0.30184	11.77164
4	61	2	63	0.03175	0.17199	10.83554
5	7	0	7	0.00000	0.11118	0.77827
6	30	5	35	0.14286	0.08650	3.02750
7	3	0	3	0.00000	0.07707	0.23122
8	30	2	32	0.06250	0.07355	2.35374
9	0	0	0		0.07225	0.00000
10	0	0	0		0.07177	0.00000
11	5	0	5	0.00000	0.07160	0.35799
12	7	1	8	0.12500	0.07153	0.57226
13	0	0	0		0.07151	0.00000
14	1	0	1	0.00000	0.07150	0.07150
15	0	0	0		0.07150	0.00000
16	14	0	14	0.00000	0.07150	1.00094
Total	212	56	268			62.14091

Table 5: n_k Values for Partial CCF Events from 1997 through 2015

Group Size	n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8
2	115.61	31.164						
3	109.23	64.693	9.267					
4	96.47	80.938	25.857	4.038				
5	96.92	73.190	39.900	14.539	2.141			
6	98.05	65.654	46.584	23.001	8.533	1.209		
7	101.15	61.106	44.388	29.499	15.401	5.219	0.716	
8	102.97	59.825	41.221	31.486	20.828	10.368	3.197	0.453

The results (i.e., the probability of complete CCF event and the estimated number of complete CCF events for each group size) are listed in the last two columns of Table 4.

Note that the binomial regression treatment of complete CCF events in Table 4 (and in Section 4) does not distinguish lethal shock events from non-lethal shock but complete CCF events. For lethal shocks, the impact vectors are supposed to map directly, i.e., the probability that all x components in a system of x components have failed due to a lethal shock is mapped directly and equals to the probability of failing all y components in a system of y components. The correct process should treat lethal shock events and the non-lethal shock but complete CCF events differently: mapping the lethal shock events directly, while curve fitting the non-lethal shock but complete CCF events.

A review of the CCF data used in this study (from 1997 through 2015) found 3 CCF events that are coded as lethal shock: 244-2005-0142, 263-1999-0046, and 423-2012-0501, all of which have a group size of 2. While the results in Table 4 are used in the following sections to estimate prior distributions, sensitivity analysis could be conducted to estimate prior distributions with different treatment of complete CCF events, i.e., using the mapping method or using the binomial regression method but distinguishing lethal shocks from non-lethal shocks.

5.3. Estimating Prior Distributions

Adjusted n_k Values

Adjusted n_k values for CCF events from 1997 through 2015 were obtained by adding the estimated number of complete CCF events in Table 4 to the final n_k value for the partial CCF events in Table 5 for each group size. For example, the estimated number of complete CCF events for group size 2 is 31.140 in Table 4; the n2 value for partial CCF events for group size of 2 in Table 5 is 31.164; the adjusted n2 value for group size of 2 will be 31.164+31.140 = 62.304. Table 6 shows the adjusted n_k results for group sizes 2 through 16 with CCF data from 1997 through 2015. The number of effective independent failure events (n_I), which is obtained from the CCF database website query results, and the total number of failures (n_t), which is the sum of the n_I and n_k, for each group size are also presented in Table 6.

Table 6: Adjusted n_k Values for CCF Events from 1997 through 2015

Group Size	n_t	n_I	n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8
2	3042.49	2864.58	115.61	62.304						
3	4491.83	4296.87	109.23	64.693	21.038					
4	5947.30	5729.16	96.47	80.938	25.857	14.873				
5	7388.92	7161.45	96.92	73.190	39.900	14.539	2.919			
6	8839.80	8593.74	98.05	65.654	46.584	23.001	8.533	4.237		
7	10283.74	10026.03	101.15	61.106	44.388	29.499	15.401	5.219	0.948	
8	11731.02	11458.32	102.97	59.825	41.221	31.486	20.828	10.368	3.197	2.807

Alpha Factor Mean Values

The MLEs or mean values of alpha factors for each group size can then be calculated with Equations 2 and 3. Table 7 presents the results.

$$\alpha_1 = \frac{n_I + n_1}{n_t} \qquad (2)$$

$$\alpha_i = \frac{n_i}{n_t} \qquad \text{for } i = 2, \ldots, m \qquad (3)$$

Table 7: Calculated Alpha Factor Mean Values for CCF Events from 1997 through 2015

Group Size	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8
2	0.9795	2.048E-02						
3	0.9809	1.440E-02	4.684E-03					
4	0.9795	1.361E-02	4.348E-03	2.501E-03				
5	0.9823	9.905E-03	5.400E-03	1.968E-03	3.951E-04			
6	0.9833	7.427E-03	5.270E-03	2.602E-03	9.653E-04	4.793E-04		
7	0.9848	5.942E-03	4.316E-03	2.869E-03	1.498E-03	5.075E-04	9.214E-05	
8	0.9855	5.100E-03	3.514E-03	2.684E-03	1.775E-03	8.838E-04	2.725E-04	2.393E-04

CalcPrior Code and Prior Distributions

Adjusted n_k values including the number of effective independent failure events (n_I) in Table 6 are used as input to the computer code CalcPrior to estimate the industry-wide prior distributions with parameters α and β. The CalcPrior code was first developed in early 2000 and then re-coded with a modern computer language for this study. The code estimates prior distributions based on the constrained noninformative and Dirichlet methodology. Table 8 shows the results from the CalcPrior code for the prior distribution parameters.

Table 8: Estimated CCF Industry-Wide Prior Distributions with CCF Events from 1997 through 2015

Group Size	a1	b1	a2	b2	a3	b3	a4	b4
2	2.2413E+01	4.6853E-01	4.6853E-01	2.2413E+01				
3	5.7979E+01	1.1280E+00	8.5122E-01	5.8255E+01	2.7682E-01	5.8830E+01		
4	9.0676E+01	1.8936E+00	1.2597E+00	9.1310E+01	4.0244E-01	9.2167E+01	2.3148E-01	9.2338E+01
5	1.9084E+02	3.4322E+00	1.9242E+00	1.9235E+02	1.0490E+00	1.9322E+02	3.8224E-01	1.9389E+02
6	2.2522E+02	3.8349E+00	1.7011E+00	2.2735E+02	1.2070E+00	2.2785E+02	5.9596E-01	2.2846E+02
7	3.7180E+02	5.7474E+00	2.2432E+00	3.7530E+02	1.6295E+00	3.7591E+02	1.0829E+00	3.7646E+02
8	3.9002E+02	5.7256E+00	2.0181E+00	3.9373E+02	1.3905E+00	3.9436E+02	1.0621E+00	3.9469E+02

Group Size	a5	b5	a6	b6	a7	b7	a8	b8
2								
3								
4								
5	7.6743E-02	1.9420E+02						
6	2.2109E-01	2.2883E+02	1.0978E-01	2.2895E+02				
7	5.6537E-01	3.7698E+02	1.9159E-01	3.7735E+02	3.4801E-02	3.7751E+02		
8	7.0259E-01	3.9505E+02	3.4974E-01	3.9540E+02	1.0784E-01	3.9564E+02	9.4689E-02	3.9565E+02

6. ISSUES/THOUGHTS ON PRIOR DISTRIBUTION DEVELOPMENT

While this paper follows the current process to develop generic prior distributions for CCF alpha factors, some issues have been noted and discussed during the study. The following provides some of the issues and preliminary thoughts on them.

(1) Is the impact vector and mapping method appropriate for CCF parameter estimations or prior determination? Are there any other alternative approaches to develop the prior distributions?

The impact vector and mapping method was introduced in NUREG/CR-4780 which was published in 1988. In order to obtain a high degree of consensus on the principles of treating CCF in risk analysis, the report has been reviewed by many experts and organizations in the U.S. and Europe. Appendix D of NUREG/CR-4780 provides a detailed discussion of the background and justification for using the mapping method in parameter estimation. While the mapping up and mapping down methods do generate a considerable amount of pseudo CCF events that are used in CCF parameter estimation, the methods seem reasonable to be used to treat scarce CCF data and estimate associated CCF parameters. Actually, the impact vector and mapping method has been used consistently in the subsequent NRC CCF studies and became the state-of-the-art in CCF event characterization and CCF parameter estimation.

In the meantime, the analyst should be aware of the uncertainty that is associated with this approach. Sensitivity analysis could be performed to understand the impact associated with this model uncertainty. Furthermore, one could always look for whether there are alternative approaches that may be better for CCF parameter estimation or prior distribution development. Actually, after nearly 30 years of using the approach, now may be a good time to collect and review what other methodologies we may have, and whether there are alternative approaches that are better suited for prior distribution development. It should be noted that whichever method is selected, there is always uncertainty related to it. So as a minimum, some kinds of sensitivity studies may be conducted to compare the mapping method with other methods to obtain prior distributions, and evaluate the impacts from different methodologies may have on the associated risk applications such as significance determination process (SDP).

(2) Is the using of binomial regression treatment of complete CCF events in the current prior development process appropriate? What should be the proper model to be used for curve fitting?

The background on why a different treatment was used for complete CCF events in the prior distribution development process is unclear, but the treatment looks like a compromise to the concern that the mapping technique is adding too many pseudo CCF events into the data analysis. Whether the treatment is appropriate and which model should be used for curve fitting may be a good topic for further discussion. Also, the current binomial regression treatment of complete CCF events does not distinguish lethal shock events from the non-lethal shock but complete CCF events. For lethal shocks, the impact vectors should be mapped directly, i.e., the probability that all x components in a system of x components have failed due to a lethal shock is mapped directly and equals to the probability of failing all y components in a system of y components. The correct process should treat lethal shock events and the non-lethal shock but complete CCF events differently: mapping the lethal shock events directly, while curve fitting the non-lethal shock but complete CCF events.

(3) How is the mapping up factor ρ determined in the current process? Is there a better way to estimate ρ?

Mapping up factor ρ is defined in the Binomial Failure Rate (BFR) model as the conditional probability of failure of each component given a non-lethal shock. It is a very important parameter in the mapping up methodology and would greatly impact the mapping results with different assumption of its value. NUREG/CR-6268 Rev.1 [8] provides a method to estimate the mapping up factor ρ with a maximum value of 0.85 being established based on the observed trends and empirical studies. In CCF Parameter Estimations 2003 Update [9], the previously recommended value of 0.85 was thought to be too conservative and 0.50 was recommended. During this study, a new process was developed to estimate the mapping up factor ρ. The process was preliminarily tested with the pump CCF data. Whether this new process to estimate ρ should be applied to the process and the CCF software is subject to further review and decision.

(4) Is there a general formula for mapping CCF data up?

A table of formulas is presented in NUREG/CR4780 (as Table D-5) [2] and NUREG/CR-5485 (as Table C-5) [7] for upward mapping of events classified as nonlethal shocks. This table is expanded in NUREG/CR-6268 Rev. 1 (as Table 7-4) [8] with the maximum size of system mapping being increased from 4 to 6. During this study, it was found that an explicit general formula for the mapping up method maybe exist. Whether this new general mapping up formula should be applied to the process and the CCF software is subject further review and decision.

(5) Should different prior distributions be developed for different component groups?

During the study, with some simple examinations of the CCF data, it was found that different component groups may have quite different alpha factors and need to be analyzed separately.

(6) Will the testing scheme for various components impact priors?

It is unclear whether different component testing scheme (staggered testing versus non-staggered or "simultaneous" testing) would impact the prior estimation, or whether separate data analyses are needed for them. One quick thought is that if, for example, some valves have staggered testing and some have nearly simultaneous testing, the data for those two kinds of valves must be analyzed separately, because the two kinds will have numerically different alphas. Mixing the two kinds of data would not give a correct result for either valve type.

7. CONCLUSION

This paper provides an overview of the history of NUREG CCF reports, documents the current process used to develop CCF prior distributions, and updates alpha factor priors with data from 1997 through 2015. The issues encountered during the CCF prior study as well as the preliminary thoughts are presented for further discussions. A list of future work is provided to address the issues listed in Section 6:

(1) Perform sensitivity analysis to understand the impact of different prior distributions on event and condition assessment. For example, identify several component groups and different prior distributions, calculate alpha parameters for each of the component groups using different priors, and plug the resultant alpha factors into one or two SPAR models.

(2) Examine the current process and methodologies to develop CCF prior distributions. For example, the general formula for mapping CCF data up, the binomial regression treatment of complete CCF events, the estimation of the mapping up factor ρ, whether different priors should be developed for different component groups for alpha factors, the calculation of the average group size in the CCF Data Software and its impact on the results, etc. Revise the potential errors in current mapping up formulas used in the CCF Data Software.

(3) Evaluate whether there are any other alternative approaches to develop prior distributions instead of the current impact vector and mapping approach.

Acknowledgements
The authors wish to thank John Lane, Chris Hunter, and Kevin Coyne of the U.S. NRC for their management support and technical inputs. Cory Atwood has been instrumental to the project by providing invaluable insights and suggestions to the study. The supports from Kellie Kvarfordt, Jim Knudsen, Ted Wood, and Mike Calley of INL are greatly appreciated. The authors would also like to pay tribute to Ali Mosleh, Karl Fleming, Dale Rasmuson, Tom Wierman, and many other analysts who make significant contribution to advance the CCF methodologies and applications.

agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the U.S. Government or any agency thereof.

References

[1] A. Mosleh, et. al., "*Procedures for Treating Common Cause Failures in Safety and Reliability Studies,"* Volume 1, United States Nuclear Regulatory Commission, NUREG/CR-4780, EPRI NP-5613, January 1988.

[2] A. Mosleh, et. al., "*Procedures for Treating Common Cause Failures in Safety and Reliability Studies,"* Volume 2, United States Nuclear Regulatory Commission, NUREG/CR-4780, EPRI NP-5613, January 1989.

[3] J. K. Vaurio, "*Common-Cause Failure Models, Data, Quantification*," Lappeenranta University of Technology, IEEE Transactions on Reliability, Vol. 48, No. 3, September 1999.

[4] F. Marshall, A. Mosleh, and D. Rasmuson, "*Common-Cause Failure Database and Analysis System,*" Volumes 1 to 3, United States Nuclear Regulatory Commission, NUREG/CR-6268, INEEL/EXT-97-00696, June 1998.

[5] K. Kvarfordt, et. al., "*Common-Cause Failure Database and Analysis System*," Volume 4, United States Nuclear Regulatory Commission, NUREG/CR-6268, INEEL/EXT-97-00696, June 1998.

[6] F. Marshall, D. Rasmuson, and A. Mosleh, "*Common-Cause Failure Parameter Estimations*," United States Nuclear Regulatory Commission, NUREG/CR-5497, INEEL/EXT-97-01328, October 1998.

[7] A. Mosleh, D. Rasmuson, and F. Marshall, "*Guidelines on Modeling Common-Cause Failures in Probabilistic Risk Assessment*," United States Nuclear Regulatory Commission, NUREG/CR-5485, INEEL/EXT-97-01327, November 1998.

[8] T. Wierman, D. Rasmuson and A. Mosleh, "*Common-Cause Failure Database and Analysis System: Event Data Collection, Classification, and Coding*," United States Nuclear Regulatory Commission, NUREG/CR-6268, Revision 1, September 2007.

[9] U.S. Nuclear Regulatory Commission, "*CCF Parameter Estimations, 2003 Update*," http://nrcoe.inel.gov/resultsdb/ParamEstSpar/, May 2006.

[10] U.S. Nuclear Regulatory Commission, "*CCF Parameter Estimations, 2005 Update*," http://nrcoe.inel.gov/resultsdb/ParamEstSpar/, September 2008.

[11] U.S. Nuclear Regulatory Commission, "*CCF Parameter Estimations, 2007 Update*," http://nrcoe.inel.gov/resultsdb/ParamEstSpar/, September 2008.

[12] U.S. Nuclear Regulatory Commission, "*CCF Parameter Estimations, 2009 Update*," http://nrcoe.inel.gov/resultsdb/ParamEstSpar/, January 2011.

[13] U.S. Nuclear Regulatory Commission, "*CCF Parameter Estimations, 2010 Update*," http://nrcoe.inel.gov/resultsdb/ParamEstSpar/, January 2012.

[14] U.S. Nuclear Regulatory Commission, "*CCF Parameter Estimations, 2012 Update*," http://nrcoe.inel.gov/resultsdb/ParamEstSpar/, November 2013.

[15] U.S. Nuclear Regulatory Commission, "*CCF Parameter Estimations, 2015 Update*," http://nrcoe.inel.gov/resultsdb/ParamEstSpar/, October 2016.

[16] C. Atwood, "*Hits per Trial: Basic Analysis of Binominal Data*," EGG-RAAM-11041, 1994.

[17] C. Atwood, et. al., "*Handbook of Parameter Estimation for Probabilistic Risk Assessment*," United States Nuclear Regulatory Commission, NUREG/CR-6823, SAND2003-3348P, September 2003.

[18] MathWorks, https://www.mathworks.com/products/matlab.html