A General Cause Based Methodology for Analysis of Common Cause and Dependent Failures in System Risk and Reliability Assessments

Andrew O'Connor\textsuperscript{a}, Ali Mosleh\textsuperscript{b}

\textsuperscript{a1} Center for Risk and Reliability, University of Maryland, College Park, United States
\textsuperscript{b2} Center for Risk and Reliability, University of Maryland, College Park, United States

Abstract: Traditional Probabilistic Risk Assessments (PRAs) model dependency through deterministic relationships in fault trees and event trees, or through empirical ratio common cause failure (CCF) models. However, popular CCF models do not recognize system specific defenses against dependencies and are restricted to identical components in redundant configuration. While this has allowed prediction of system reliability with little or no data, it is a limiting factor in many applications, such as modeling the characteristics of a system design or incorporating the characteristics of failure when assessing the failure’s risk significance or degraded performance events (known as an event assessment).

This paper proposes the General Dependency Model (GDM), which uses Bayesian Network to model the probabilistic dependencies between components. This is done through the introduction of three parameters for each failure cause which relate to physical attributes of the system being modelled, component fragility, cause condition probability, and coupling factor strength.

Finally this paper demonstrates the development and use of the GDM for new system PSA applications and event assessments of existing system. Examples of the quantification of the GDM model in the presence of uncertain evidence are provided.

Keywords: Common Cause Failure, Bayesian Network, General Dependency Model, Dependency Modelling.

1. INTRODUCTION

Probabilistic system modeling aims to understand system vulnerabilities, compare alternative designs and estimate exposure to risk in steady state and scenario based conditions. Traditionally system models only included discrete dependencies between components, however, greater knowledge of how systems interact and respond to events has enabled explicit treatment of ‘soft’ dependencies within systems such as a common manufacturer or technician through Common Cause Failure models. Without proper treatment of these dependencies, the system safety estimate can be severely underestimated and the system vulnerabilities within specific systems are not well understood.

An example of soft dependencies causing unexpected system failure is Eastern Air Lines Flight 855, where the failure of all three engines was caused by a loss of oil from missing O-ring seals from each engine. The NTSB identified the probable cause as “failure of mechanics to follow the established and proper procedures for the installation of master chip detectors in the engine lubrication system, the repeated failure of supervisory personnel … and the failure of Eastern Air Lines management…” [1] The failure of each engine cannot be treated as independent as they share maintenance crews, supervisors and management system.

Common Cause Failure (CCF) models either apply a qualitative or quantitative approach. Qualitative models involve a qualitative assessment of the soft dependencies between components and apply generic factors to adjust the Probability Safety Assessment (PSA) results. This provides the ability to better understand the vulnerabilities of the system and implement defenses to protect against CCFs. These estimators used in the PSA are usually subjective and not based on collected data. Quantitative models

1 Current affiliation, Relken Engineering Pty Ltd, Australia
2 Current affiliation, University of California, Los Angeles, USA
use data from previously observed events to quantitatively assess the impact on the probability of system failure. These methods typically use impact vectors as described in NUREG/CR-5497 [2] which apply assumptions of symmetry between components and cannot quantify the effect of system specific defenses against CCF. Despite previous attempts, it has been difficult to combine the cause information from qualitative assessments with data used on quantitative methods. [3,4] Further, while these models allow for the quantification of soft dependencies at a system level, they are inadequate for providing further insight into the causes of CCF, which is necessary to inform mitigation strategy development.

The limitations of the current CCF models is particularly apparent when assessing a particular failure’s risk significance within the system. Without being able to incorporate the specific failure cause and propensity for propagation through the system given that system’s defenses, the assessment is limited in its insights of the system’s response. It has become evident that the commonly accepted CCF modeling methodology [2] and corresponding tools need enhancements to meet these PSA activities.

This paper proposes the General Dependency Model (GDM), which uses a Bayesian Network to model the probabilistic dependencies between components. GDM allows the explicit modelling of system specific dependencies and defenses and can be implemented across non-identical components. The model

2. DESCRIBING COMMON CAUSE FAILURES

CCF requires two factors; a failure cause and a coupling factor[2]. Failure cause is the condition that the component failure can be attributed to, the coupling factor is the dependency between components which propagates the failure to multiple components. A Common Cause Component Group (CCCG) is a group of components which share coupling factors, making them susceptible to a common failure cause while defenses are the parts of a system that protect against the failure cause or the coupling factor.

CCFs are only a problem when failures occur within a timeframe that multiple components cannot provide their function. This is sometimes called ‘simultaneous failure.’ It should be noted that the term ‘simultaneous’ is relative and may be defined using a mission period or a replacement/repair period.

3. CURRENT CCF ANALYSIS METHODOLOGY AND ITS LIMITATIONS

Broadly, CCF models fit into the following categories:

- **Basic Parameter Model** – calculates the CCF basic event directly from data [5]. The Basic Parameter model cannot estimate CCFs for redundancy configurations for which data is unavailable, and for this reason is rarely used directly.
- **Ratio Models** (e.g.Beta Factor, Alpha Model and Multiple Greek Letter Models) – assume the number of CCF events is a transferable empirical ratio between failure rates and CCF rate [6]. Ratio estimates are calculated from generic data from sources like the NRC CCF Data Base (CCFDB) [7] and combined with system specific failure rates to obtain CCF estimates.
- **Shock Models** – assume that each component within the CCCG receive shocks according to a Poisson process. Each failure event is modelled from a Bernoulli trial of each component to the shock.

All models rely on each component belonging to only one CCCG. This forces the assumption that each component within the CCCG is identical in design, but also identical in its dependencies between each other component within the CCCG. Dependencies to components outside the CCCG cannot be modelled. This restricts CCF modelling to only identical components in redundant configurations with identical dependencies, which is rarely accurate.

Furthermore, all models rely on impact vectors from generic CCF databases. Impact vectors do not account for the type of failure cause or coupling factors. These models cannot incorporate system specific defenses against coupling factors or causes, and the system response to a specific failure cause cannot be assessed.
4. GENERAL DEPENDENCY MODEL (GDM)

The GDM has been proposed to enable event assessment with knowledge of the failure event’s characteristics. It can model the increased and decreased propensity of a system to experience CCF based on the system features such as causes, coupling factors and defenses and seeks to model asymmetrical components and dependency relationships. The GDM will retain the modeling of different multiplicities of failures and allow for parameter estimation using the impact vector methodology.

4.1. Model Structure

4.1.1. Component Failure Probability

The GDM defines the component failure rate, \( Q_t \), as the combination of component failure probabilities for each failure cause. To the component, each cause is independent of each other and the component failure probability can be calculated using rare event approximation:

\[
P(A) = Q_t = \sum_{i=1}^{w} Q_{t,i}
\]

\( A = \) A random variable for the failure of component \( A \)

\( Q_t = \) The total failure probability for a component

\( Q_{t,i} = \) The failure probability of a component due to cause \( i \).

\( C_i \), is a condition from which a failure can occur due to cause \( i \). The probability that a cause condition exists is \( P(C_i) = Q_{E,i} \). Given \( C_i \), the component failure probability is \( p_i \), the probability of component failure due to cause \( i \) is:

\[
P_{t,i} = p_i Q_{E,i}
\]

\[
p_i = \text{the probability a component fails when tested by cause } i.
\]

4.1.2. Component Dependency

Thus far the model has included the cause conditions and failure probabilities which are local to a component. It is possible that multiple components share the same cause condition due to a coupling factor. The presence of coupling factors between components is identified during the qualitative assessment of the target system features as described in NUREG/CR-5497 [2]. By coupling the cause condition, instead of failures, the model can better describe the physical phenomena of CCF and model asymmetrical relationships.
For example, the figure below shows an emergency diesel generator (EDG) and pump which share the same location. If the EDG suffers from an extreme environmental condition, then the pump will also experience the same condition (or shock). The difference between the EDG and pump in the presence of such a cause condition, is the fragility of each component to withstand the shock, $p_i$.

![Figure 2: GDM Coupling Components](image)

4.1.3. Propagation of Cause Condition

The model, thus far, has assumed a certain (deterministic) propagation of a cause condition to other components, where a coupling factor exists. However the propagation of a cause needs to be probabilistic. For example, an inexperienced tradesman makes a maintenance error on an EDG before progressing to maintain a second EDG. Despite the maintainer coupling the two components, the likelihood of the second EDG suffering the same maintenance error is probabilistic. Without making this relationship the model could not account for occasions when components have high fragility to a cause condition, and a low probability of CCF or where defenses against cause condition propagation exist such as protecting against environmental causes by moving components into separate rooms.

The GDM separates the local cause condition for each component. Local cause conditions can propagate to other components probabilistically using a coupling strength factor, $\eta_i$.

The coupling factor strength, $\eta_i$, needs to scale between the following two extremes. When the coupling factor strength is zero, $\eta_i = 0$, there is no chance that the local cause condition at one component can propagate to the second component. When the coupling factor strength is one, $\eta_i = 1$, a cause condition can only be present at both components simultaneously. GDM splits each local cause into cause condition probability, $Q_{E,i}$, into independent ($Q_{IE,i}$) and common error ($Q_{CE,i}$) probabilities:

$$ P(X_i) = Q_{CE,i} = \eta_i Q_{E,i} $$
$$ P(I_i) = Q_{IE,i} = (1 - \eta_i) Q_{E,i} $$

$X_i$ = random variable for the common cause condition for cause i.
$I_i$ = random variable for the independent cause condition for cause i.
$\eta_i$ = the coupling factor strength for cause i.

The common and independent cause conditions are mutually exclusive events. Therefore the local cause condition probability is the sum of the independent and common cause condition probabilities.

$$ C_i = I_i \cup X_i $$
$$ Q_{E,EE} = Q_{IE,EE} + Q_{CE,EE} $$

$C_i$ = A random variable for the existence of cause condition i.
$Q_{E,i}$ = The cause condition probability of cause i.

Figure 3 shows the construction of the GDM with consideration for a coupling factor strength parameter.
4.1.4. Parameter Description

One strength of GDM is that the parameters can be interpreted as physical features of the system being modelled. For each failure cause classification, the GDM has three parameters, the component fragility, the cause condition probability, and the coupling factor strength.

**Fragility.** $p_i$ is the probability a component will fail given that a cause condition is evident for cause $i$. It is a measure of the components ability to resist failure. The component’s fragility is affected by such things as the component’s design, materials, derating and compliance to reliability durability standards.

**Cause Condition Probability.** $Q_{E,i}$ is the probability that the Cause condition for cause $i$ is present. The cause condition probability represents the frequency and strength of failure causes. It is a function of features such as quality assurance, process maturity, and human performance shaping factors. This parameter is similar to the Binomial Failure Rate Model’s, rate of shocks and max exist for extended periods of time.

**Coupling Factor Strength.** $\eta_i$ is the probability that if a cause condition exists at a component, that it will be propagated to other components. The coupling factor strength is a measure of the strength in dependency between components including coupling factor defenses.

4.2. Parameter Estimation

For each cause the model is fully specified once the three parameters $p_i, Q_{E,i}, \eta_i$ are known. However, using data from the NRC failure databases, the observable quantities are:

- The failure rate for a component due to cause $i, Q_{t,i}$.
- The propensity for CCF due to cause $i$ in a perfectly symmetrical CCCG, $\alpha_{2,i}$.

4.2.1. GDM Relationship to $Q_{t,i}$

The failure probability for cause $i$, $Q_{t,i}$ is an observable metric which can assist in the calculation of the GDM parameters through the relationship:

$$Q_{t,i} = p_i Q_{E,i}$$  \hspace{1cm} (7)
The point estimate for the failure probability of a component due to cause \( i \), \( Q_{t,i} \) is:

\[
Q_{t,i} = \frac{n_{F,i}}{N_i}
\]

(8)

\( n_{F,i} = \) the total number of failures due to cause \( i \).
\( N_i = \) the total number of demands on a single component.

The quantities, \( n_{F,i} \) and \( N_i \) are component event data, as opposed to CCF event data. For example, if two components are in redundancy, assume that when the system is demanded, both components are demanded. In the first demand, component 1 fails. In the second demand, component 2 fails, in the third demand, no components fail. Then each component was demanded 3 times, making a total of 6 component demands for the system with two failures. The failure probability is \( Q_{t,i} = \frac{2}{6} = \frac{1}{3} \).

4.2.2. GDM Relationship to \( \alpha_{z,i} \)

The CCF data measures the strength of a coupling factor through the frequency of CCF events observed. In the AFM, this is quantitatively measured through the use of alpha factors. An alpha factor specific to each cause, (Partial Alpha Factor) may be calculated using impact vectors [8] and therefore \( \alpha_{z,i} \) it is a convenient measure to use for estimating GDM parameters.

GDM uses an assumption that each component has a Bernoulli trial in the presence of a cause condition, and will fail with probability \( p_i \). Therefore higher multiplicities of failure are not explicitly modeled, which is similar in concept to the Binomial Failure Rate Model (BFRM). Furthermore the assumption required to estimate the AFM parameters requires all components within the CCCG were perfectly symmetrical in design, use and dependencies. The higher the size of a CCCG, the less likely it is this assumption is satisfied. Therefore it is proposed that only the second alpha factor is required. This has the advantage that there is likely to be much more data on CCCGs with two components, than larger groups.

In order to obtain a relationship between \( \alpha_{z,i} \) and GDM, the results from the event assessment of a two train, perfect symmetry system has been analyzed using both the Partial Alpha Factor method [8] and GDM. Using Partial Alpha Factors the probability of failure for a component \( A \) given knowledge of component B failing due to cause \( i \), is:

\[
P(A|B) = \alpha_{z,i}^2 Q_t + \alpha_{z,i}
\]

(9)

The \( \alpha_2 \) term is the probability of CCF while the remaining term, \( \alpha_{z,i}^2 Q_t \), is the normalized probability of independent failure for component \( A \). The same calculation can be done using GDM. The cause node for component B, \( C_i^{[B]} \), is instantiated as true, and the probability for the second component A failing due to that cause is calculated as:

\[
P \left( A_i \left| C_i^{[B]} \right. \right) = \frac{p_i Q_{E,i}(\eta_i - 1)^2}{1 - Q_{E,i}(\eta_i - 1)} + p_i \eta_i
\]

(10)

It can be shown that the term \( p_i \eta_i \) is the probability of CCF. The remaining term, is the normalized probability of independent failure for component

Therefore, when the GDM model is calculated for two components with perfect symmetry:

\[
p_i \eta_i = \alpha_{z,i}
\]

(11)

When the failure cause taxonomy is defined in such a way that each cause could only propagate through
one coupling factor, the point estimate for the partial alpha factor is:

\[ \alpha_{z,i} = \frac{n_{z,i}}{n_{t,i}} \]  

(12)

Where

\begin{align*}
\alpha_{z,i} & = \text{a partial alpha factor which represents the portion of system failure events which resulted in 2 components failing within a common cause component group of size 2 when there was a potential for failure propagation through coupling factor } i \text{ where } i \in \{1,2,3,...,w\} \\
n_{k,i} & = \text{the number of failure events/frequency which resulted in } k \text{ components failing within a common cause component group of size } m, (1 \leq k \leq 2) \text{ of coupling factor } i \text{ where } i \in \{1,2,3,...,w\} \\
n_{t,i} & = \text{the total number of CCF events for coupling factor/cause } i \text{ where } i \in \{1,2,3,...,w\}.
\end{align*}

4.2.3. Estimation Using Observed Data

The two relationships have been established, \( Q_{t,i} = p_i Q_{E,i} \) and \( \alpha_{z,i} = p_i \eta_i \) above. To complete the estimation of the GDM parameters, one of the three parameters must be estimated through other means. Quantification of the remaining parameter can be conducted from:

- Direct assessment from data which represents a parameter.
- Using constraints from asymmetrical components.
- Assume \( \eta_i = 1 \) as per Binomial Failure Rate Model.
- Estimate from parametric failure model, such as human reliability models for human cause conditions, \( Q_{E,i} \) or load strength interference model for \( p_i \).
- Engineering assessment.
- Solve using data from higher levels of alpha factors.

Where a components share a cause, a third constraint is imposed where the cause condition rate of both components must be equal. The third equation is:

\[ Q_{CE,i} = \eta_i Q_{E,i} \]  

(14)

Each type of cause lends itself better to different types of estimation techniques for the third parameter for example:

- **Human Error Cause.** The area of Human Reliability Assessments is rich with literature and parametric models. Therefore human causes may be best suited to parametric modeling to estimate \( Q_{E,i} \) or \( \eta_i \). Failing this, an engineering assessment of \( \eta_i \) would be the next best option.
- **Procedural Error Cause.** Where components are coupled by the same procedure, it is highly likely that if one component is affected, then all shared components may be affected. Therefore procedural errors may be suitable for the assumption \( \eta_i = 1 \) or an expert elicitation estimate of \( \eta_i \).
- **Environmental Error Cause.** Unlike the other causes, environmental cause conditions may be detectable using sensors. In such cases, \( Q_{E,i} \) may be estimated directly from cause condition data. Where this is not possible, the propagation of environmental causes will change between systems depending on the location and building design housing components. Therefore environmental causes may be suitable for an expert elicitation estimate of \( \eta_i \).
4.3. GDM in Event Assessment

A strength of the GDM model is the flexibility when conducting event assessments. Three event assessment scenarios will be presented:

- Event assessment with knowledge of a component failure
- Event assessment with knowledge of a component failure and failure cause
- Event assessment with virtual evidence about the component failure cause

In order to demonstrate the model capabilities within event assessment an example involving non-asymmetrical components will be used.

4.3.1. Example: Two Train EDG and Pump System

The example system consists of a mixture of pumps and generators with varying levels of dependency. The systems objective is to provide water to a cooling system. The two train system, only requires one train to be running in order to provide sufficient water. A pump requires power from an Emergency Diesel Generator to operate. One of the trains has two pumps in redundancy, resulting in a total of three pumps for the system.

The failure probability for an EDG is also assumed to be $Q_{t}^{E} = 0.006$ and the failure probability for a pump is assumed to be $Q_{t}^{P} = 0.00204$. The reliability block diagram is shown in Figure 4.

![Figure 4: Reliability block diagram - Two EDGs and three pump system](image)

P(s) is $4.82e^{-5}$ and the cut sets are: \( \{E_1, E_2\} ; \{P_1, P_2, P_3\} ; \{E_1, P_2, P_3\} \)

4.3.2. Qualitative Analysis

<table>
<thead>
<tr>
<th>Component</th>
<th>Install Procedure</th>
<th>Maintenance Staff</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDG 1 (E1)</td>
<td>EDG</td>
<td>Team X</td>
<td>Room Y</td>
</tr>
<tr>
<td>EDG 2 (E2)</td>
<td>EDG</td>
<td>Team X</td>
<td>Room Y</td>
</tr>
<tr>
<td>Pump 1 (P1)</td>
<td>Pump V1.1</td>
<td>Team X</td>
<td>Room Y</td>
</tr>
<tr>
<td>Pump 2 (P2)</td>
<td>Pump V2.8</td>
<td>Team X</td>
<td>Room Y</td>
</tr>
<tr>
<td>Pump 3 (P3)</td>
<td>Pump V1.1</td>
<td>Team Y</td>
<td>Room X</td>
</tr>
</tbody>
</table>

Figure 5 shows the GDM structure for the example system with local cause conditions being removed for brevity. All dependencies can be represented within the model without the assumption of symmetry. Quantification of this model will be discussed in a future paper.
4.3.3. Knowledge of Failure

The procedure for conducting event assessment using the Bayesian Network is shown in Figure 9. The analyst applies evidence to the node and the other node values are updated. The system failure probability has increased from $2.152 \times 10^{-4}$ to $3.810 \times 10^{-2}$ with only knowledge of the pump failure. Notice that the failure probabilities of the EDGs increases because the pump failure may have been caused by the maintenance team or external environment which is shared by the EDGs.

4.3.4. Knowledge of Failure Cause

Where the failure cause is known, the system equation can be updated by instantiating the local cause node as true. In the example an event assessment where Pump 1 has failed due to a Maintenance Human cause condition is shown in Figure 7. The EDG1, EDG2, Pump 1, and Pump 2 all share the same maintenance team. With Pump 1 failing due to a human error from that maintenance team, the Bayesian Network now propagates this evidence, and increases our belief that EDG1, EDG2 and Pump 2 could fail. The probability of system failure with this new evidence increases from $3.810 \times 10^{-2}$ to $5.265 \times 10^{-2}$. 
The probability of system failure for each possible cause for Pump 1 is contained in Table 3. As can be seen, if the cause had of been an installation error, the system failure probability would have dropped from 3.810e-2 to 1.351e-2. This would have been due to Pump 1 only sharing the same installation procedure with Pump 3 and so the options for the failure to propagate to the rest of the system is limited.

Table 3: Event Assessment for Example 2 with different failure causes

<table>
<thead>
<tr>
<th>Cause</th>
<th>$P(S)$</th>
<th>GDM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unknown</td>
<td>$P(S</td>
<td>A)$</td>
</tr>
<tr>
<td>Install Procedure Error</td>
<td>$P(S</td>
<td>A,C_{IP})$</td>
</tr>
<tr>
<td>Maintenance Human Error</td>
<td>$P(S</td>
<td>A,C_{MH})$</td>
</tr>
<tr>
<td>External Environment Shock</td>
<td>$P(S</td>
<td>A,C_{EE})$</td>
</tr>
</tbody>
</table>

4.3.5. Uncertain Knowledge of Failure Cause

CCF characteristics may be difficult to interpret from reports. The impact methodology aims to capture this uncertainty in a CCF database, however this cannot be used when conducting event assessment. Virtual Evidence can be applied to the Bayesian Network where the analyst is unsure of a node’s state. Such evidence during event assessment might be available from an initial incident investigation where the failure cause cannot distinguish between causes. The analyst must estimate the odds that each node is true over the other node states.

For example, an analyst might believes there is a 30:70 chance that the failure of pump 1 was due to cause Maintenance Human, and a 70:30 chance that it was due to Installation Procedure. The external environment cause has been ruled out. The Bayesian Network before and after the virtual evidence has been applied is shown in Figure 8 and Figure 9 respectively. The probability that the failure was caused by Maintenance Human has dropped from 0.71 to 0.32. Given the more likely cause is now a pump installation problem, the system probability has dropped from 0.04 to 0.02.
4.4. Extensions and Future Development of GDM

This paper has focused on a model to replace current component level CCF models. However, there are a number of possible extensions to improve the model’s accuracy, flexibility or maturity, including:

- **Scalability** – Use of a Bayesian Network, allows the GDM to model multiple levels of causality through either the GDM cause condition construct, or normal Bayesian Network nodes to model probabilistic dependencies. The definition of cause condition is not confined to a level of causality and is easily redefined. However, an alternative approach is to create a network of nodes between the cause condition and the component failure.
• **Consistency of Asymmetrical Components** – While the GDM is already capable of asymmetrical component modelling, improvements can be made to the procedures for ensuring consistency, particularly when multiple asymmetrical components require integration.

• **Failure Taxonomy Development** – Current NRC CCF taxonomy is ambiguous when inferring which failure causes could propagate through which coupling factors. An industry informed, unambiguous failure classification taxonomy for CCF will be necessary.

5. **CONCLUSION**

Traditional PRAs model dependency through deterministic relationships in fault trees and event trees or through empirical ratio CCF (CCF) models. Current CCF models are restricted to identical components in redundant formations and cannot incorporate system specific defenses against dependencies. Due to these restrictions current CCF models are not suitable for event assessments.

To overcome these limitations, this paper proposed the General Dependency Model (GDM), which uses Bayesian Networks to model the probabilistic dependencies between components via three parameters for each failure, fragility, cause condition probability, and coupling factor strength which relate to physical attributes of the system being modelled.

Through the use of an example, this paper showed how an analyst could build and quantify the GDM using similar inputs to that of current CCF methodologies. It also showed how the GDM can accommodate uncertain evidence, asymmetrical components, coupling factor strength, as well as providing insight into the change in propensity of a system to experience CCF based on actual system features.

6. **ACKNOWLEDGEMENTS**

This work was funded through Collaborative Research Grant NRC-04-10-164 by the U.S. Nuclear Regulatory Commission (USNRC). The authors thank the NRC Project Technical Monitor, Dr. Jeffrey Wood for his interest and support throughout the project. The authors also appreciate valuable technical guidance and feedback offered by Dr. Song Hua Shen, Dr. Kevin Coyne, Mr. Gary Demoss, Dr. Nathan Siu, and Mr. Don Marksberry, from the NRC Office of Research. While this report was prepared under a collaborative research grant and benefitted from extensive discussions with the NRC staff, USNRC has neither approved nor disapproved its technical content.

7. **REFERENCES**


