Theoretical comparison of models for a seismically induced joint failure probability

Shota Soga[a], Eishiro Higo[b], and Hiromichi Miura[c]
[a] Central Research Institute of Electric Power Industry, Yokosuka, Japan, soga@criepi.denken.or.jp
[b] higo3791@criepi.denken.or.jp
[c] h-miura@criepi.denken.or.jp

Abstract: An earthquake simultaneously challenges multiple structures, systems, and components of a nuclear power plant. Seismic probabilistic risk assessment evaluates this phenomenon with a failure condition that a component fails when a seismic response exceeds a component capacity. In literature, there are several models for a seismically induced joint failure probability: a model used in the seismic safety margins research program (Model 1), a model in the SECOM2 (Model 2), and the Reed-McCann procedure (Model 3). We also discuss a model that applies the separation of independent and common variables method to response and capacity (Model 4). In Model 4, common variables among more than two components are explicitly considered. These four models are analytically compared to clarify their relationship with respect to correlation coefficients. First, it is shown that the first two models are equivalent by showing their derivations. However, Model 1 is advantageous because there are efficient algorithms to evaluate it. Next, Model 4 is shown as a limited case of Model 1 and 2 using a characteristic function. Finally, Model 3 is shown as a limited case of Model 4 by deriving the failure criterion used in Model 3 from Model 4 by neglecting common variables among more than two components. Thus, we summarize the relation: Model 1 = Model 2 ⊃ Model 4 ⊃ Model 3 with respect to correlation coefficients. Therefore, we recommend Model 1 for a joint failure probability because of its computational efficiency and better applicability.

1. INTRODUCTION

An earthquake is considered one of the major risk contributors to a nuclear power plant. Unlike an internal event such as a turbine trip initiating event caused by random failure, an earthquake is unique because it simultaneously affects multiple structures, systems, and components (SSCs). Therefore, researchers have proposed several models for a joint probability of seismically induced component failures [1–3]. The seismic safety margin research program (SSMRP) [2] introduced a model which calculates the probability that seismic responses of multiple components simultaneously exceed their capacities (Model 1). This model results in an orthant probability, and there is an efficient algorithm to evaluate this probability [4]. Then, a different model was proposed by SECOM2 [3,5] (Model 2). After the SSMRP model, Reed et al. proposed the so-called Reed-McCann procedure [1] (Model 3), in which a failure criterion of a component is defined in terms of ground motion (GM). Furthermore, experts on seismic fragility analysis have published NUREG/CR-7237, which recommends the separation of independent and common variables (SICV) method* [6]. It is important to note that the detailed derivations of these models have not been provided in the literature, making it challenging to analyze and compare these models analytically.

Therefore, this paper aims to provide detailed derivations of these models and compare them analytically. We show the equivalence of Model 1 and Model 2 by showing that they are different parameterizations of the same probability distributions. In addition, we discuss a model developed by applying the SICV method to response and capacity (Model 4). We show that Model 4 is a limited case of Model 1. Furthermore, model 4 can only consider nonnegative correlations, whereas Model 1 can

* The Reed-McCann procedure is also known as the separation of independent and common variables (SICV) method because the Reed-McCann procedure is the SICV method applied to ground motion capacity and median GM capacity.

also consider all possible correlations. Finally, we show that Model 3 is a limited case of Model 4. In summary, we reveal the relation of the existing models: Model 1 = Model 2 ⊃ Model 4 ⊃ Model 3 in correlation coefficients.

2. COMPARISON OF THE EXISTING MODELS

We compare the existing models for computing a joint failure probability. Before deriving these models, let us introduce notations. Let n denote the number of components and subscript i and j denote the indexes of them. First, let $S_i(A)$ and T_i denote response and capacity of ith component. The response is a function of peak ground acceleration (PGA) A. Bolded variables denote vectors, matrices, and sets. Thus, $\boldsymbol{S}(A)$ and $\boldsymbol{T}$ denote vectors of responses and capacities, respectively. Throughout this paper, we omit (A) if it is obvious. We assume that $\boldsymbol{S}$ and $\boldsymbol{T}$ are assumed multivariate lognormal distributions written as $\boldsymbol{S}(A) \sim \mathcal{MLN}(\boldsymbol{\mu_S}(A), \boldsymbol{\Sigma_S}(A))$ and $\boldsymbol{T} \sim \mathcal{MLN}(\boldsymbol{\mu_T}, \boldsymbol{\Sigma_T})$ where $\boldsymbol{\mu}$ is a vector of the logarithm means and $\boldsymbol{\Sigma}$ is the covariance matrix whose ijth elements are $\Sigma_{S,ij} = \beta_{S,i}\beta_{S,j}\rho_{\ln S_i, \ln S_j}$ and $\Sigma_{T,ij} = \beta_{T,i}\beta_{T,j}\rho_{\ln T_i, \ln T_j}$. $\beta_{S,i}$ and $\beta_{T,i}$ are the logarithmic standard deviations of ith response and capacity, respectively. $\rho_{x,y}$ is the correlation coefficient of the pair of random variables $\{x, y\}$. For example, $\rho_{\ln S_i, \ln S_j}$ is the correlation coefficient between $\ln S_i$ and $\ln S_j$.

We also use $\mathcal{N}$ for normal distribution and $\mathcal{LN}$ for a lognormal distribution. Let $\beta_{C,i}$ denote composite uncertainty of ith component defined as $\beta_{C,i}^2 = \beta_{S,i}^2 + \beta_{T,i}^2$. Response and capacity are also lognormally distributed as $S_i \sim \mathcal{LN}(\mu_{S,i}, \beta_{S,i})$ and $T_i \sim \mathcal{LN}(\mu_{T,i}, \beta_{T,i})$, respectively, where $\mu_{S,i}$ and $\mu_{T,i}$ are ith element of $\boldsymbol{\mu_S}$ and $\boldsymbol{\mu_T}$ as described above. A median of a lognormal distribution $\mathcal{LN}(\mu, \sigma)$ is $\exp(\mu)$. Therefore, if we assume a linear response [5] in which response ith component is written as $w_i A$, there exists a PGA value $A_{m,i}$ such that the medians of response and capacity of ith component are equal. That is, $w_i A_{m,i} = \exp\left(\mu_{S,i}(A_{m,i})\right) = \exp(\mu_{T,i})$ where $\mu_{S,i}$ is the ith element of $\boldsymbol{\mu_S}$. If a response is nonlinear, then w_i is a function of A as $w_i(A)$. We use G for GM capacity to distinguish it from PGA A, where GM capacity is defined as the PGA value for which the seismic response of a given component exceeds the component capacity [7], and GM capacity has uncertainty. Throughout this paper, G has uncertainty, and A does not. This distinction aims to adapt the concept in the Reed-McCann procedure that assumes uncertainty in GM capacity and median GM capacity. Also, $G_{m,i}$ denotes the median GM capacity of ith component.

Figure 1 summarizes the assumptions and steps in the derivations of models discussed in this paper. Table 1 summarizes the existing models for obtaining a joint failure probability P. All models have a common ground: the failure criterion and the lognormality assumption. In seismic PRA, we assume that a component fails when a response exceeds capacity. In application, engineers use a failure criterion in terms of PGA such that a component fails when PGA exceeds GM capacity.

Table 1: Equations for a joint failure probability. *P* denotes a joint failure probability of components (1/2).

Model	Equation
Model 1 (SSMRP [2])	**Failure criterion**: Response > Capacity **Joint failure probability**: $$P = \frac{1}{(2\pi)^{n/2}\lvert\boldsymbol{V}_1\rvert^{1/2}} \int_0^\infty \cdots \int_0^\infty \exp\left(-\frac{1}{2}(\boldsymbol{z} - \boldsymbol{\mu_z})^{\mathrm{T}} \boldsymbol{V_z}^{-1} (\boldsymbol{z} - \boldsymbol{\mu_z})\right) \mathrm{d}\boldsymbol{z}$$ where $$\boldsymbol{z} = \ln \boldsymbol{S} - \ln \boldsymbol{T}$$ $$\boldsymbol{\mu_z} = \boldsymbol{\mu_S} - \boldsymbol{\mu_T}$$ $$\boldsymbol{V_z} = \boldsymbol{\Sigma_S} + \boldsymbol{\Sigma_T}$$

Table 1: Equations for a joint failure probability. P denotes a joint failure probability of components (2/2).

Model	Equation
Model 2 (JAERI [3])	**Failure criterion**: Response > Capacity **Joint failure probability**: $$P=\frac{1}{(2\pi)^{n/2}\lvert\boldsymbol{V}_2\rvert^{1/2}}\int_{-\infty}^{u_1}\cdots\int_{-\infty}^{u_n}\exp\left(-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{V}_2^{-1}\boldsymbol{x}\right)\mathrm{d}\boldsymbol{x}$$ where $$x_i=\frac{\mu_{S,i}-\mu_{T,i}-\ln S_i+\ln T_i}{\beta_{C,i}}\dagger$$ $$u_i=\frac{\mu_{S,i}-\mu_{T,i}}{\beta_{C,i}}$$ $$\boldsymbol{V}_2=\boldsymbol{b}(\boldsymbol{\Sigma}_{\boldsymbol{S}}+\boldsymbol{\Sigma}_{\boldsymbol{T}})\boldsymbol{b}^{\mathrm{T}}$$ $$\boldsymbol{b}=\mathrm{diag}\left(1/\beta_{C,1},\cdots,1/\beta_{C,n}\right)$$
Model 3 (Reed-McCann [1])	**Failure criterion**: PGA > GM capacity **Joint failure probability**: OR condition (at least one component fails): $$P=\int_0^\infty \mathrm{d}\boldsymbol{G}'_{\boldsymbol{m}}\left[h(\boldsymbol{G}'_{\boldsymbol{m}})\int_0^\infty \mathrm{d}\boldsymbol{G}'\left\{g(\boldsymbol{G}')\left(1-\prod_{i=1}^{n}(1-f_i)\right)\right\}\right]$$ AND condition (all components fail): $$P=\int_0^\infty \mathrm{d}\boldsymbol{G}'_{\boldsymbol{m}}\left[h(\boldsymbol{G}'_{\boldsymbol{m}})\int_0^\infty \mathrm{d}\boldsymbol{G}'\left\{g(\boldsymbol{G}')\prod_{i=1}^{n}f_i\right\}\right]$$ where $$f_i=\Phi\left(\frac{\ln\frac{A}{G'_{m,ii}\prod_{i=1,i\neq j}^{n}G'_{ij}G'_{m,ij}}}{\beta_{ii}}\right)$$ $$g(\boldsymbol{G}')=\prod_{i=1}^{n-1}\prod_{j=i+1}^{n}\frac{1}{\beta_{ij}G'_{ij}}\psi\left(\frac{\ln G'_{ij}}{\beta_{ij}}\right),\text{ and}$$ $$h(\boldsymbol{G}'_{\boldsymbol{m}})=\prod_{i=1}^{n}\prod_{j=i}^{n}\frac{1}{\beta_{m,ij}G'_{m,ij}}\psi\left(\frac{\ln G'_{m,ij}}{\beta_{m,ij}}\right)$$ $\boldsymbol{G}'$ is the set of G'_{ij} except G'_{ii}. $\boldsymbol{G}'_{\boldsymbol{m}}$ is set of $G'_{m,ij}$. See Section 2.3 for the definitions of G'_{ij} and $G'_{m,ij}$.

† In the original paper, x_i is expressed as $\ln(T_i/S_i)$. Since x_i is the variable of integration, the joint failure probability is the same.

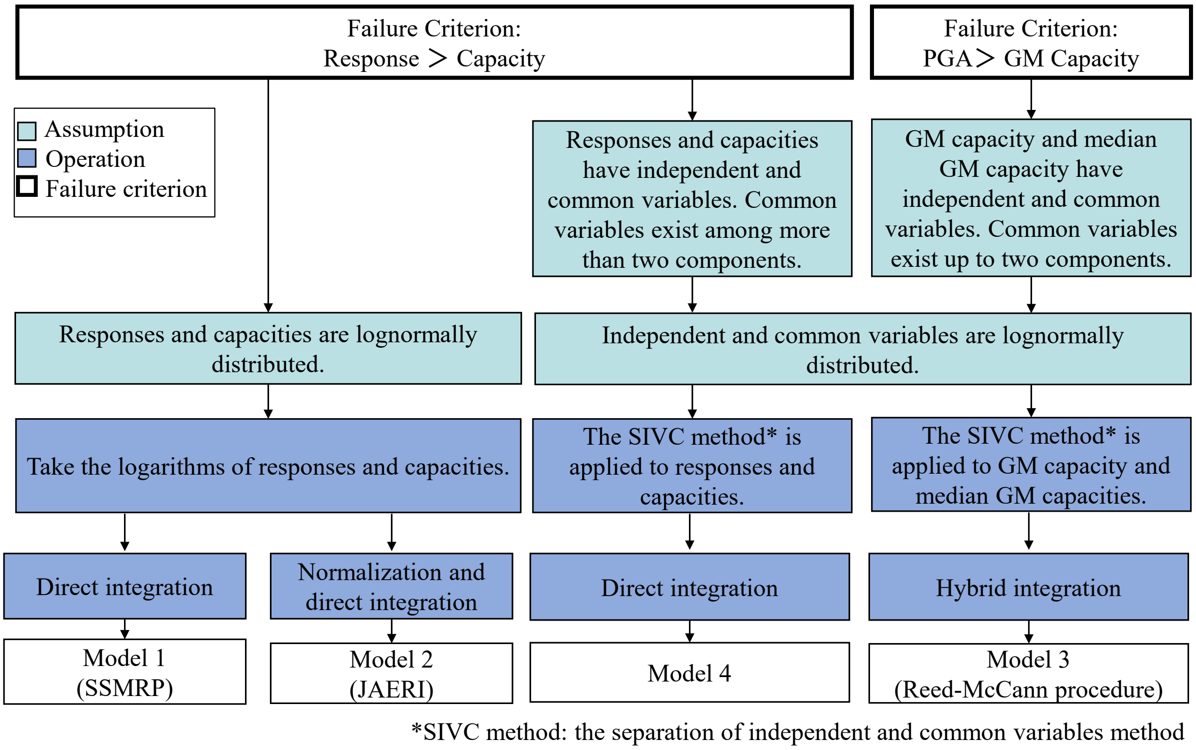

Figure 1: Important assumptions and steps in derivations
Note that one can derive the failure criterion: PGA>GM capacity from the other failure criterion: Response > Capacity, as described in Section 2.5.

2.1. Derivation of the SSMPR model (Model 1)

This model was originally proposed in the SSMRP report [2]. The failure criterion is transformed into a joint failure probability step by step. In this model, the failure criterion for ith component, $S_i > T_i$, is first transformed into the equivalent failure criterion, $\ln(S_i) - \ln(T_i) > 0$ by taking logarithms. From the definition of the lognormal distribution, $\ln(S_i)$ and $\ln(T_i)$ are normally distributed, and the subtraction of two normal variables also follows a normal distribution. Thus, $\ln(S_i) - \ln(T_i) \sim \mathcal{N}\left(\mu_{s,i} - \mu_{T,i}, \beta_{T,i}^2 + \beta_{S,i}^2\right)$. Similarly, the failure criterion that all components fail simultaneously can be written as $\ln(\boldsymbol{S}) - \ln(\boldsymbol{T}) > \boldsymbol{0}$[‡], and $\ln(\boldsymbol{S}) - \ln(\boldsymbol{T})$ follows a multivariate normal distribution written as $\ln(\boldsymbol{S}) - \ln(\boldsymbol{T}) \sim \mathcal{MVN}(\boldsymbol{\mu_S} - \boldsymbol{\mu_T}, \boldsymbol{\Sigma_T} + \boldsymbol{\Sigma_S})$. Let $\boldsymbol{z}$ equal $\ln(\boldsymbol{S}) - \ln(\boldsymbol{T})$, and its probability density function (pdf) is expressed as

$$\mathrm{pdf}(\boldsymbol{z}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{V}_1|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{z} - \boldsymbol{\mu_z})^{\mathrm{T}} \boldsymbol{V}_1^{-1} (\boldsymbol{z} - \boldsymbol{\mu_z})\right)$$

where $\boldsymbol{V}_1 = \boldsymbol{\Sigma_T} + \boldsymbol{\Sigma_S}$ and $\boldsymbol{\mu_z} = \boldsymbol{\mu_S} - \boldsymbol{\mu_T}$. Hence, the probability that all components satisfy the failure criterion is the volume over $\boldsymbol{z} > \boldsymbol{0}$. Accordingly, the joint failure probability is given as

$$P(\boldsymbol{z} > \boldsymbol{0}) = \int_0^\infty \cdots \int_0^\infty \mathrm{pdf}(\boldsymbol{z})\, \mathrm{d}\boldsymbol{z}$$

This probability is also known as an orthant probability because it evaluates one orthant of a probability space. The advantage of Model 1 is that there exist efficient algorithms to evaluate this probability and software (for example, reference [4] and mvn.cdf function in SciPy.stats packages [8]). Note that there is no restriction in correlation coefficient values. Therefore, they can be any value bound in $[-1,1]$.

[‡] $\ln(\boldsymbol{S})$ means that the vector of natural logarithms of all elements of $\boldsymbol{S}$.

Model 1 does not use the linear response assumption. If we assume the linear response, $\boldsymbol{\mu}_z$ is expressed as

$$\boldsymbol{\mu}_z = \boldsymbol{\mu}_{\boldsymbol{S}} - \boldsymbol{\mu}_{\boldsymbol{T}} = \begin{bmatrix} \ln(w_1 A) \\ \vdots \\ \ln(w_n A) \end{bmatrix} - \begin{bmatrix} \ln(w_1 A_{m,1}) \\ \vdots \\ \ln(w_n A_{m,n}) \end{bmatrix} = \begin{bmatrix} \ln(A) - \ln(A_{m,1}) \\ \vdots \\ \ln(A) - \ln(A_{m,n}) \end{bmatrix}$$

where w_i is the linear coefficient for the ith component.

2.2. Derivation of the JAERI model (Model 2)

This model was originally described in the SECOM2 manual [3]. We show that Model 2 is equivalent to Method 1 by showing that both models are based on the same failure criterion, the same assumption, and different parameterizations. In Model 2, the failure criterion is transformed by normalization as

$$\begin{aligned} S_i > T_i &\Leftrightarrow 0 > \ln(T_i) - \ln(S_i) \\ &\Leftrightarrow -(\mu_{T,i} - \mu_{S,i}) > \ln(T_i) - \ln(S_i) - (\mu_{T,i} - \mu_{S,i}) \\ &\Leftrightarrow \underbrace{\frac{\mu_{S,i} - \mu_{T,i}}{\beta_{C,i}}}_{u_i} > \underbrace{\frac{\ln(T_i) - \ln(S_i) + \mu_{S,i} - \mu_{T,i}}{\beta_{C,i}}}_{x_i}, \end{aligned}$$

where x_i follows a standard normal distribution. A failure is now written as $u_i > x_i$. Similarly, the multivariate lognormal distributions can be transformed into the multivariate normal distribution with zero means and correlation matrix as a covariance matrix as $\boldsymbol{x} \sim \mathcal{MVN}(0, \boldsymbol{V}_2)$, where $\boldsymbol{V}_2 = \boldsymbol{b}(\boldsymbol{\Sigma}_{\boldsymbol{S}} + \boldsymbol{\Sigma}_{\boldsymbol{T}})\boldsymbol{b}$ and $\boldsymbol{b}$ is a diagonal matrix with its ith diagonal element equal to $1/\beta_{C,i}$. Thus, the pdf of $\boldsymbol{x}$ is given as

$$\mathrm{pdf}(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{V}_2|^{1/2}} \exp\left(-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{V}_2^{-1}\boldsymbol{x}\right),$$

and the joint failure probability is given as

$$P(\boldsymbol{u} > \boldsymbol{x}) = \int_{-\infty}^{u_1} \cdots \int_{-\infty}^{u_n} \mathrm{pdf}(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}.$$

Model 2 starts from the same assumptions as Model 1 but normalizes its failure criterion for each component, $\ln(T_i) - \ln(S_i)$ whereas Model 1 does not normalize $\ln(S_i) - \ln(T_i)$. This normalization does not change a probability value if these models use the same covariance matrices and means. Thus, we obtain Model 1 = Model 2.

2.3. Derivation of the Reed-McCann procedure (Model 3)

Model 3 was originally proposed by Reed et al. [1]. The original Reed-McCann procedure is the two-step procedure that incorporates numerical and analytical integrations. The first step applies the SICV method to median GM capacity values. Then, these values are sampled using the Latin hypercube method. In the second step, the joint failure probability is estimated using the sampled median GM capacity values and uncertainty of GM capacities. We also show that Model 3 is a limited case of Model 1 in Section 2.5 by showing Model 1⊃Model 4 ⊃Model 3.

Model 3 is based on the SICV method applied to the median GM capacity and GM capacity. In the SICV method, a variable is decomposed into an independent variable and common variables, where each common variable is shared among two or more variables. The original Reed-McCann procedure only considers the common variables shared in possible combinations of two variables. It also assumes that these variables are lognormally distributed, and the logarithmic means of common variables equal zero.

We derive an equation for the failure probability of ith component. First, let us assume that G_i and $G_{m,i}$ denote the GM capacity and median GM capacity of the ith component, and β_i and $\beta_{m,i}$ are logarithmic standard deviations of G_i and $G_{m,i}$ respectively. Both capacities follow a lognormal distribution. Then, we use the property of a lognormal distribution. Assume a random variable x following a lognormal distribution expressed as $x \sim \mathcal{LN}(\mu, \beta)$. Then, let $\mathbb{M}[x]$ denote the median of x and equal $\exp(\mu)$. Now, let us assume that $x' \sim \mathcal{LN}(0, \beta)$, and we multiply x' by a scalar value a, resulting in a new lognormal distribution $ax' \sim \mathcal{LN}(\ln a, \beta)$. If $a = \exp(\mu) = \mathbb{M}[x]$, ax' is equal to x. Thus, a lognormal distribution can be expressed as the product of its median and a random variable following a lognormal distribution with zero logarithmic mean. Now, we apply this property to G_i resulting

$$G_i = G_{m,i} G_i', \tag{1}$$

where $G_{m,i} \sim \mathcal{LN}(\mu_i, \beta_{m,i})$ and $G_i' \sim \mathcal{LN}(0, \beta_i)$. Note that the Reed-McCann procedure assumes an uncertainty in the median GM capacity, so we assumed $G_{m,i}$ is lognormally distributed. Once again, we apply the same property to $G_{m,i}$ resulting

$$G_{m,i} = \exp(\mu_i)\, G_{m,i}', \tag{2}$$

where $G_{m,i}' \sim \mathcal{LN}(0, \beta_{m,i})$. Note that μ_i has no uncertainty, so $\exp(\mu_i) = A_{m,i}$. Thus,

$$G_{m,i} = A_{m,i} G_{m,i}'. \tag{3}$$

The Reed-McCann procedure samples a set of $G_{m,i}$ by the Latin hypercube sampling method. Now, Model 3 applies the SICV method to G_i' and $G_{m,i}'$:

$$G_i' = \prod_{j=1}^{n} G_{ij}' \tag{4}$$

$$G_{m,i}' = \prod_{j=1}^{n} G_{m,ij}', \tag{5}$$

where G_{ii}' and $G_{m,ii}'$ are independent variables and G_{ij}' and $G_{m,ij}'$ for $i \neq j$ are common variables between ith and jth components. Note that we assume that these variables are statistically independent. This assumption is necessary to derive Eqs. (8) and (11). From the assumption of the SICV method that G_{ij}' and $G_{m,ij}'$ are lognormally distributed, expressed as $G_{ij}' \sim \mathcal{LN}(0, \beta_{ij})$ and $G_{m,ij}' \sim \mathcal{LN}(0, \beta_{m,ij})$, these uncertainties have identities:

$$\beta_i^2 = \sum_{j=1}^{n} \beta_{ij}^2$$

$$\beta_{m,i}^2 = \sum_{j=1}^{n} \beta_{m,ij}^2.$$

Then, Model 3 assumes a failure criterion that an ith component fails when PGA exceeds GM capacity, written as $A > G_i$. The failure criterion for the ith component is transformed as

$$\begin{aligned} P(A > G_i) &= P(A > G_{m,i} G_i') = P\left(\frac{A}{G_{m,i} \prod_{i=1, i\neq j}^{n} G_{ij}'} > G_{ii}'\right) \\ &= P\left(\frac{\ln \frac{A}{G_{m,i} \prod_{i=1, i\neq j}^{n} G_{ij}'}}{\beta_{ii}} > \underbrace{\frac{\ln G_{ii}'}{\beta_{ii}}}_{z(\text{normalization})}\right) = \Phi\left(\frac{\ln \frac{A}{G_{m,i} \prod_{i=1, i\neq j}^{n} G_{ij}'}}{\beta_{ii}}\right) \end{aligned} \tag{6}$$

where $\Phi(x)$ is the cumulative distribution function of a standard normal distribution. Thus, the probability that all components fail is the product of conditional probabilities as $\prod_{i=1}^{n} P(A > G_i)$.

G_{ij}' is a random variable. Hence, we can integrate it out from Eq.(6). Let $\varphi(x, \mu, \beta)$ denote a pdf of a lognormal distribution where x is the random variable, μ is the logarithmic mean, and β is the logarithmic standard deviation. So, the probability density function of G_{ij} is expressed as

$$\varphi(G_{ij}', 0, \beta_{ij}) = \frac{1}{G_{ij}' \sqrt{2\pi\beta_{ij}^2}} \exp\left(-\frac{(\ln G_{ij}')^2}{2\beta_{ij}^2}\right) = \frac{1}{G_{ij}'\beta_{ij}} \psi\left(\frac{\ln G_{ij}'}{\beta_{ij}}\right) \tag{7}$$

where $\psi(x)$ is the probability density function of a standard normal distribution. Now, multiplying Eqs. (6) and (7) and integrating G_{ij}' result in the joint failure probability given $G_{m,i}$ values as

$$\int_0^\infty \cdots \int_0^\infty g(\boldsymbol{G}') \prod_{i=1}^{n} \Phi\left(\frac{\ln \frac{A}{G_{m,i} \prod_{i=1, i\neq j}^{n} G'_{ij}}}{\beta_{ii}}\right) \mathrm{d}\boldsymbol{G}' \tag{8}$$

where $\boldsymbol{G}'$ is the set of G'_{ij} excluding G'_{ii} and $g(\boldsymbol{G})$ is the product of $\varphi(G'_{ij}, 0, \beta_{ij})$ written as

$$g(\boldsymbol{G}') = \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} \frac{1}{\beta_{ij} G'_{ij}} \psi\left(\frac{\ln G'_{ij}}{\beta_{ij}}\right). \tag{9}$$

This integral equation agrees with the formula shown in Appendix B of [9]. The original Reed-McCann procedure samples $G_{m,i}$ using the Latin hypercube sampling, but one can also integrate it analytically. Then, using Eq.(5), the joint pdf of $G'_{m,i}$ is expressed as

$$h(\boldsymbol{G'_m}) = \prod_{i=1}^{n} \prod_{j=i}^{n} \frac{1}{\beta_{m,ij} G'_{m,ij}} \psi\left(\frac{\ln G'_{m,ij}}{\beta_{m,ij}}\right) \tag{10}$$

where $\boldsymbol{G'_m}$ is the set of $G'_{m,ij}$ including $j = i$. Substituting Eq.(3) and Eq.(5) into Eq.(8), multiplying it by (10), and integrating it over $\boldsymbol{G'_m}$ gives the joint probability as

$$P = \int_0^\infty \mathrm{d}\boldsymbol{G'_m} \left[h(\boldsymbol{G'_m}) \int_0^\infty \mathrm{d}\boldsymbol{G}' \left\{ g(\boldsymbol{G}') \prod_{i=1}^{n} \Phi\left(\frac{\ln \frac{A}{A_{m,i} G'_{m,ii} \prod_{i=1, i\neq j}^{n} G'_{ij} G'_{m,ij}}}{\beta_{ii}}\right)\right\}\right]. \tag{11}$$

Eq.(11) is the complete integral form of the Reed-McCann procedure, which is not explicitly shown in the original paper. The difference is that Eq.(11) analytically integrates $\boldsymbol{G'_m}$ whereas the original Reed-McCann procedure numerically integrate $\boldsymbol{G'_m}$. One of the obvious observations is that the number of integrals increases as the number of components increases. Therefore, the Reed-McCann procedure gets computationally intractable even with the small number of components.

Reed et al. [1] claimed that the method can consider the uncertainty in response and capacity. However, we did not consider the uncertainty in response in the derivation. One can argue that either GM or median GM capacities include uncertainty in response. However, this idea is not clearly shown in the derivation. Hence, there is a gap between the failure criterion used in Model 1 and Model 3. Later in Subsection 2.5, we analytically derive the equivalence condition of these criteria.

2.4. SICV method applied to responses and capacities (Model 4)

The original Reed-McCann procedure is the SICV method applied to GM capacities and median GM capacities. In this section, we apply the SICV method to response and capacity. We do not derive the integral form of Model 4 because Model 4 is a limited case of Model 1.

Let us assume that response and capacity decompose into independent and common variables. Unlike the Reed-McCann procedure, we consider common variables among more than two components. First, let us introduce some set notations to simplify the summation symbols. Let a number denote a component identification, and numbers in a pair of curly brackets indicate a set of components. For example, S_2 means response of the second component, including independent and common variables whereas $S_{\{2\}}$ and $S_{\{2,3\}}$ is an independent variable unique to the second component and a common variable between components 2 and 3, respectively. Let $\boldsymbol{E}$ and i denote a set of components and an element of $\boldsymbol{E}$, respectively. Then, let $s(\boldsymbol{E}, i)$ denote a set of subsets of $\boldsymbol{E}$ containing i defined as $\{\boldsymbol{x} \subseteq \boldsymbol{E} | \, \boldsymbol{x} \cap \{i\} \neq \emptyset\}$. For example, if $\boldsymbol{E} = \{1,2,3\}$ **and** $i = 1$, then $s(\{1,2,3\}, 1)$ is $\{\{1\}, \{1,3\}, \{1,2,3\}\}$**.** This notation represents responses as

$$S_i = \prod_{\boldsymbol{k} \in s(\boldsymbol{E}, i)} S_{\boldsymbol{k}} \Leftrightarrow \ln S_i = \sum_{\boldsymbol{k} \in s(\boldsymbol{E}, i)} \ln S_{\boldsymbol{k}} \tag{12}$$

Now, $\ln S_{\boldsymbol{k}}$ follows a normal distribution. Let us assume that the number of elements in a set $\boldsymbol{x}$ is expressed by $|\boldsymbol{x}|$. For example, $|\boldsymbol{k}| = 1$ means only one element in $\boldsymbol{k}$, so $S_{\boldsymbol{k}}$ represents an independent variable of a single component. If $|\boldsymbol{k}| > 1$, then $S_{\boldsymbol{k}}$ represents a common variable among $\boldsymbol{k}$. Following the argument in the Reed-McCann procedure, we can assume the logarithmic mean of $\ln S_{\boldsymbol{k}}$ for $|\boldsymbol{k}| > 1$ is zero. Eq.(12) can be written as a matrix form as

$$\underbrace{\begin{bmatrix} \ln S_1 \\ \ln S_2 \\ \vdots \\ \ln S_n \end{bmatrix}}_{\ln \boldsymbol{S}} = \underbrace{\begin{bmatrix} 1 & 0 & \cdots & 0 & 1 & 1 & \cdots & 1 \\ 0 & 1 & \cdots & 0 & 1 & 0 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & 1 & 0 & 0 & \cdots & 1 \end{bmatrix}}_{\boldsymbol{M}} \underbrace{\begin{bmatrix} \ln S_{\{1\}} \\ \ln S_{\{2\}} \\ \vdots \\ \ln S_{\{n\}} \\ \ln S_{\{1,2\}} \\ \ln S_{\{1,3\}} \\ \vdots \\ \ln S_{\{1,2,\cdots,n\}} \end{bmatrix}}_{\boldsymbol{x}}$$

where the matrix $\boldsymbol{M}$ can be interpreted as a transformation matrix from decomposed independent and common variables to original variables. For example, if $\boldsymbol{E} = \{1,2,3\}$, the above equation is

$$\begin{bmatrix} \ln S_1 \\ \ln S_2 \\ \ln S_3 \end{bmatrix} \underbrace{=}_{\text{Eq.(12)}} \begin{bmatrix} \ln S_{\{1\}} & & & +\ln S_{\{1,2\}} & +\ln S_{\{1,3\}} & & +\ln S_{\{1,2,3\}} \\ & \ln S_{\{2\}} & & +\ln S_{\{1,2\}} & & +\ln S_{\{2,3\}} & +\ln S_{\{1,2,3\}} \\ & & \ln S_{\{3\}} & & +\ln S_{\{1,3\}} & +\ln S_{\{2,3\}} & +\ln S_{\{1,2,3\}} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \ln S_{\{1\}} \\ \ln S_{\{2\}} \\ \ln S_{\{3\}} \\ \ln S_{\{1,2\}} \\ \ln S_{\{1,3\}} \\ \ln S_{\{2,3\}} \\ \ln S_{\{1,2,3\}} \end{bmatrix}.$$

The rank of M is n because M can be partitioned into $\boldsymbol{M} = [\boldsymbol{I}_n \quad \boldsymbol{M}_{n\times(2^n-n-1)}]$, and columns of $\boldsymbol{M}$ are linearly independent. $\boldsymbol{x}$ is a multivariate normal distribution with zero correlation. Now, we show that $\ln \boldsymbol{S}$ is a multivariate normal distribution with mean $\boldsymbol{M\mu}$ and covariance matrix $\boldsymbol{M\Sigma M}^T$ using a characteristic function.

The sketch of the proof is the following. Let $\boldsymbol{\mu}_x$ and $\boldsymbol{\Sigma}_x$ denote the logarithmic mean and the covariance matrix of $\boldsymbol{x}$, respectively. Then, the characteristic function of $\boldsymbol{x}$ is given as

$$\varphi_X(\boldsymbol{t}) = \exp\left(i\boldsymbol{\mu}_x^{\mathrm{T}}\boldsymbol{t} - \frac{1}{2}\boldsymbol{t}^T\boldsymbol{\Sigma}_x\boldsymbol{t}\right)$$

where i is the imaginary number, and $\boldsymbol{t}$ is the vector of real values. Now, we can express the characteristic function of S as

$$\varphi_{\ln \boldsymbol{S}}(\boldsymbol{t}) = \exp\left(i\boldsymbol{\mu}_x^{\mathrm{T}}\boldsymbol{M}^T\boldsymbol{t} - \frac{1}{2}\boldsymbol{t}^T\boldsymbol{M}\boldsymbol{\Sigma}_x\boldsymbol{M}^T\boldsymbol{t}\right).$$

The rank of $\boldsymbol{\Sigma}_x$ and $\boldsymbol{M}$ is $2^n - n$ and n, respectively, where $2^n - n \geq n$. Hence, the rank of $\boldsymbol{M\Sigma}_x\boldsymbol{M}^T$ is also n for $n \geq 2$. This guarantees that there exists an inverse matrix of $\boldsymbol{M\Sigma}_x\boldsymbol{M}^T$. Now, $\varphi_{\ln \boldsymbol{S}}(\boldsymbol{t})$ is the same functional form as a multivariate normal distribution, so, $\ln \boldsymbol{S}$ is a multivariate normal distribution with mean $\boldsymbol{M\mu}$ and covariance matrix $\boldsymbol{M\Sigma M}^T$. Thus, $\boldsymbol{S}$ is the multivariate lognormal distribution. The same argument is valid for capacity. Note that all elements of $\boldsymbol{M\Sigma M}^T$ are nonnegative, so correlation coefficients are also nonnegative because all elements of $\boldsymbol{M}$ and $\boldsymbol{\Sigma}_x$ are nonnegative. In other words, Model 4 only considers positive correlations. Note that this positiveness of $\boldsymbol{M}$ stems from how we decomposed $\boldsymbol{S}$. To consider negative correlations, one needs to include more common variables that affect responses asymmetrically, such as a common variable that increases component A's response but decreases B's response. However, these common variables are not explicitly considered in the SICV method, so we did not include them in this paper.

The above argument states that response and capacity follow multivariate lognormal distributions if the SICV method is applied. This result means that engineers do not need to estimate the common variables but correlation coefficients of the possible combinations of two components, reducing the number of required parameters. Thus, the resultant distribution is equivalent to the assumption of Models 1 and 2, except that correlation coefficients are limited to be positive. Thus, Model 4 is a limited case of Models 1 and 2. Therefore, we get the relation: Model 1 = Model 2 ⊃ Model 4.

One of the insights obtained in this derivation is that Models 1 and 2 can consider common variables among more than two components. The above result states that correlation coefficients hold information on common variables shared among three or more components.

2.5 Relation with the Reed-McCann procedure

The Reed-McCann procedure is derived from the SICV method applied to GM and median GM capacities. We show that the Reed-McCann procedure is a limited case of Model 4. The sketch of the proof is the following. We first show that the Reed-McCann procedure is expressed as the product of random variables. Then, Model 4 is reduced to this product, neglecting higher-order terms.

Using Eq.(1)-(5), one can show the relation,

$$G_i = A_{m,i} \prod_{j=1}^{n} G'_{ij} G'_{m,ij}.$$

Thus, the failure criterion in Model 3 is written as

$$A > A_{m,i} \prod_{j=1}^{n} G'_{ij} G'_{m,ij} \tag{13}$$

Now, we transform the other failure criterion into this criterion in the following manner. First, we apply the SICV method to the response and capacity of ith component as

$$S_i > T_i \Leftrightarrow \prod_{\boldsymbol{i} \in s(\boldsymbol{E},i)} S_{\boldsymbol{i}} > \prod_{\boldsymbol{i} \in s(\boldsymbol{E},i)} T_{\boldsymbol{i}}.$$

In the Reed-McCann procedure, common variables among more than two components are not considered, so we neglect common variables more than two components as

$$S_i > T_i \Rightarrow \prod_{j=1}^{n} S_{ij} > \prod_{j=1}^{n} T_{ij} \tag{14}$$

where S_{ii} and T_{ii} are independent variables, and S_{ij} and T_{ij} for $i \neq j$ are common variables. Let us assume that logarithmic standard deviations of S_{ij} and T_{ij} are written as $\beta_{S,ij}$ and $\beta_{T,ij}$ respectively.

Now, following the same argument in Eq.(3), S_{ii} and T_{ii} are expressed as $S_{ii} = wAS'_{ii}$ and $T_{ii} = wA_{m,i}T'_{ii}$ where $S'_{ii} \sim \mathcal{LN}(0, \beta_{S,ii})$ and $T'_{ii} \sim \mathcal{LN}(0, \beta_{T,ii})$. Then, substituting these into Eq.(14) results in

$$\begin{aligned} S_i > T_i &\Rightarrow S_{ii} \prod_{j=1, j\neq i}^{n} S_{ij} > T_{ii} \prod_{j=1, j\neq i}^{n} T_{ij} \\ &\Leftrightarrow wAS'_{ii} \prod_{j=1, j\neq i}^{n} S_{ij} > wA_{m,i}T'_{ii} \prod_{j=1, j\neq i}^{n} T_{ij} \\ &\Leftrightarrow AS'_{ii} \prod_{j=1, j\neq i}^{n} S_{ij} > A_{m,i}T'_{ii} \prod_{j=1, j\neq i}^{n} T_{ij} \end{aligned} \tag{15}$$

Now, we use a different property of a lognormal distribution again. Let us assume a random variable $x \sim \mathcal{LN}(0, \beta)$. Now, let us consider its inverse $1/x$. The logarithm of $1/x$ is equal to $\ln 1/x = -\ln x$. Since x and $1/x$ are lognormally distributed, its logarithm is normally distributed. Then, the normal distribution with zero mean is symmetric with respect to the origin. Thus, the probability of $1/x$ is equal to that of x if they have the same realization value. This property preserves the probability of inequality after replacing S_{ij} with $1/S_{ij}$. Thus,

$$P(S_i > T_i) \approx P\left(AS'_{ii}\prod_{j=1,j\neq i}^{n} S_{ij} > A_{m,i}T'_{ii}\prod_{j=1,j\neq i}^{n} T_{ij}\right)$$
$$= P\left(A > A_{m,i}T'_{ii}S'_{ii}\prod_{j=1,j\neq i}^{n} S_{ij}T_{ij}\right) \quad (16)$$

Now, we got the same form as the failure criterion used by the Reed-McCann procedure. Eq.(16) is an approximation because it neglects higher-order terms. If we do not neglect them, it becomes equivalent. These random variables have logarithmic means equal to zero, the same as the Reed-McCann procedure. The necessary condition that those failure criteria are equivalent is that all terms are equal in Eq.(13) and Eq.(16). That is,

$$T'_{ii}S'_{ii} = G'_{ii}G'_{m,ii}, \text{ and}$$
$$S_{ij}T_{ij} = G'_{ij}G'_{m,ij}.$$

Both sides of the equations are lognormal distributions with zero logarithmic means. Thus, these equalities hold when the following equality holds:

$$\beta_{S,ij}^2 + \beta_{T,ij}^2 = \beta_{ij}^2 + \beta_{m,ij}^2. \quad (17)$$

Eq.(17) tell us how to transform seismic and capacity uncertainties into aleatory and epistemic uncertainty without altering a joint failure probability.

Now we show the sketch of the proof of Model 4 ⊃ Model 3 using the proof by contradiction. Suppose Eq.(17) does not hold. In that case, Eq.(13) no longer represents the failure criterion of a component that its response is greater than its capacity. Therefore, Eq.(17) must hold so that Model 3 represents a component failure. When Eq.(17) holds, Model 4 is equivalent to Model 3 by neglecting common variables among more than two components. This shows the relation: Model 4 ⊃ Model 3.

NUREGF/CR-7237 described β_{ij} and $\beta_{m,ij}$ as aleatory and epistemic uncertainties, respectively. So, let us decompose $\beta_{S,ij}$ and $\beta_{T,ij}$ into aleatory (R) and epistemic (U) uncertainties with subscripts R and U. That is,

$$\beta_{S,ij}^2 = \beta_{S,ij,R}^2 + \beta_{S,ij,U}^2$$
$$\beta_{T,ij}^2 = \beta_{T,ij,R}^2 + \beta_{T,ij,U}^2$$

Then, these uncertainties are organized into aleatory and epistemic parts, resulting in

$$\beta_{ij}^2 = \beta_{S,ij,R}^2 + \beta_{T,ij,R}^2 \quad (18)$$
$$\beta_{m,ij}^2 = \beta_{S,ij,U}^2 + \beta_{T,ij,U}^2. \quad (19)$$

Eq.(18) and Eq.(19) provide one interpretation of β_{ij} and $\beta_{m,ij}$. Note that there are many other interpretations of β_{ij} and $\beta_{m,ij}$. For example, $\beta_{ij} = \beta_{S,ij}$ and $\beta_{m,ij} = \beta_{T,ij}$ satisfy Eq.(17), and the joint failure probability is the same as the conditions Eq.(18) and (19).

It is important to note that it is not shown whether Model 4 ⊈ Model 3 holds or not, so we show Model 4 ⊈ Model 3. There are combinations of β_{ij} and $\beta_{m,ij}$ such that they result in the same value of correlation coefficients from Method 4. However, to achieve this, we must ease the assumption in the SICV method that independent and common variables are statistically independent. If those variables are not statistically independent, the derivation of Model 3 does not hold. Thus, Model 4 ⊈ Model 3.

3. CONCLUSION

We derived the models for a joint probability of seismically induced component failures. The derivation revealed the relation of these models with respect to correlation coefficients: Model 1 = Model 2⊃ Model 4 ⊃ Model 3. Model 1 and Model 2 can consider positive and negative correlation coefficients. However, Model 4 can consider only the positive correlation coefficients, and Model 3 is the limited case of Model 4. Therefore, we recommend Model 1 (the SSMRP model) because it is the most applicable model and there is an efficient generic algorithm for it.

References

[1] J. W. Reed, M. W. McCann, J. Iihara, et al. "*Analytical techniques for performing probabilistic seismic risk assessment of nuclear power plants*", International conference on structural safety and reliability; Kobe (Japan); 27-29 May 1985.

[2] P. D. Smith, R.G. Dong, D.L. Bernreuter, et al. NUREG/CR-2015: Seismic safety margins research program: Phase I: Final Report-Overview. Washington, DC; 1981.

[3] T. Uchiyama, T. Oikawa, M. Kondo, et al. User's Manual of SECOM2: A Computer Code for Seismic System Reliability Analysis (in Japanese), JAERI-DATA/CODE--2002-011, 2002.

[4] A. Genz, F. Bretz. Computation of multivariate normal and t probabilities. Springer Science & Business Media; 2009.

[5] Subcommittee of External Hazard PRA. A Standard for Procedure of Seismic Probabilistic Risk Assessment for Nuclear Power Plants: 2015 [in Japanese]. Atomic Energy Society of Japan; 2015.

[6] R. J. Budnitz, G.S. Hardy, D.L. Moore, et al. Correlation of Seismic Performance in Similar SSCs (Structures, Systems, and Components). US Nuclear Regulatory Commission. 2017.

[7] R. P. Kennedy, C. A. Cornell, R.D. Campbell, et al. "*Probabilistic seismic safety study of an existing nuclear power plant*", Nuclear Engineering and Design. 1980;59.

[8] E. Jones, T. Oliphant, P. Peterson, et al. SciPy: Open Source Scientific Tools for Python, 2001 (http://www.scipy.org/). Http://Www.Scipy.Org/. 2015.

[9] W.S. Jung, K. Hwang, S.K. Park. A new methodology for modeling explicit seismic common cause failures for seismic multi-unit probabilistic safety assessment. Nuclear Engineering and Technology. 2020;52.