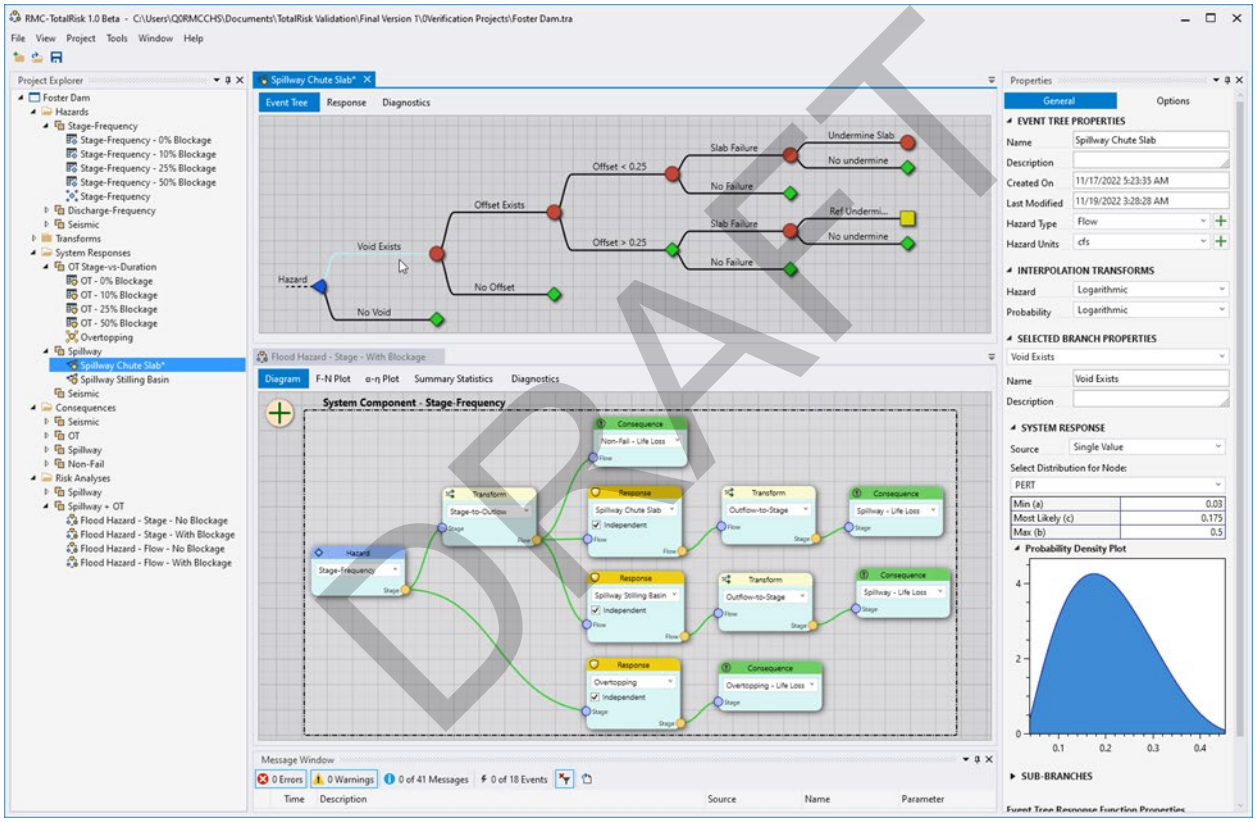
Quantitative Risk Analysis with the RMC-TotalRisk Software

RMC-TR-2023-XX



US Army Corps of Engineers®
Institute for Water Resources
Risk Management Center

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
	Technical Report	

4. TITLE AND SUBTITLE

Quantitative Risk Analysis with the RMC-TotalRisk Software

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

C. Haden Smith, Risk Management Center, U.S. Army Corps of Engineers
Woodrow L. Fields, Risk Management Center, U.S. Army Corps of Engineers
David A. Margo, Risk Management Center, U.S. Army Corps of Engineers

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Risk Management Center
12596 West Bayaud Ave. Suite 400
Lakewood, CO 80228

8. PERFORMING ORGANIZATION REPORT NUMBER

RMC-TR-2023-XX

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
RMC-TR-2023-XX

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The U.S. Army Corps of Engineers TotalRisk software (RMC-TotalRisk) performs quantitative risk calculations for a system from a set of hazard, system response, and consequence functions. RMC-TotalRisk is part of an integrated software suite designed to support risk assessments. The software features a fully integrated modeling platform, including a modern graphical user interface, data entry, report quality charts, and diagnostic tools. RMC-TotalRisk can evaluate risk for a single component with multiple failure modes and a complex system comprised of multiple components. This document provides a technical reference for the mathematical framework, numerical methods, and computation algorithms used in RMC-TotalRisk.

15. SUBJECT TERMS

Software, quantitative risk analysis, probability distributions, uncertainty analysis, numerical methods, dam safety, levee safety, flood risk management, planning, infrastructure

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	UU	87	

Quantitative Risk Analysis with the RMC-TotalRisk Software

July 2023

U.S. Army Corps of Engineers
Institute for Water Resources
Risk Management Center
12596 West Bayaud Ave. Suite 400
Lakewood, CO 80228

Author(s)

C. Haden Smith, USACE Risk Management Center

Woodrow L. Fields, USACE Risk Management Center

David A. Margo, USACE Risk Management Center

Contents

Figures

Tables

This page blank intentionally

Acknowledgements

RMC-TotalRisk would not exist without the support of Risk Management Center (RMC) leadership, in particular the RMC Director Nathan Snorteland and Lead Engineers, Jason Needham, and Timothy O'Leary. The authors would like to acknowledge Jesse Morrill-Winter and Dean McLeod for providing background material and support from the Planning Community of Practice within USACE. The authors would like to also recognize David Bowles and Ruben Jongejan for performing an external peer review of the software and risk analysis framework. The software development team is very grateful to those who helped contribute to the software.

Introduction

The U.S. Army Corps of Engineers (USACE) Risk Management Center (RMC) developed the quantitative risk analysis software (RMC-TotalRisk) to enhance and expedite risk assessments within the USACE Flood Risk Management, Planning, and Dam and Levee Safety Communities of Practice.

RMC-TotalRisk is a menu-driven software, which performs risk analysis from user defined hazard, system response, and consequence functions. The software features a fully integrated modelling platform, including a modern graphical user interface, data entry capabilities, and report quality charts and diagnostics. TotalRisk can perform multi-failure risk analysis for a single dam or levee or for a complex system with multiple components.

This document provides a technical reference for performing quantitative risk analysis with RMC-TotalRisk. The mathematical details, numerical methods, and simulation algorithms underlying each method are presented. Enough technical detail is provided so that the reader can replicate the methods if desired.

This document is aimed at engineers and scientists that have studied probability and statistics at a college-level. It is recommended that the reader refresh themselves on probability fundamentals in Appendix A before continuing.

Examples related to flood risk for dams and levees are provided throughout this document. However, the software is not limited to just flood risk management applications. RMC-TotalRisk is a general-purpose risk analysis software, capable of estimating risk for a variety of complex systems.

RMC Software Suite

The RMC-TotalRisk software is part of the greater RMC risk analysis software suite [1]. Figure 1 shows a schematic of the software suite and how each tool is intended to interact together in support of the overall risk analysis.

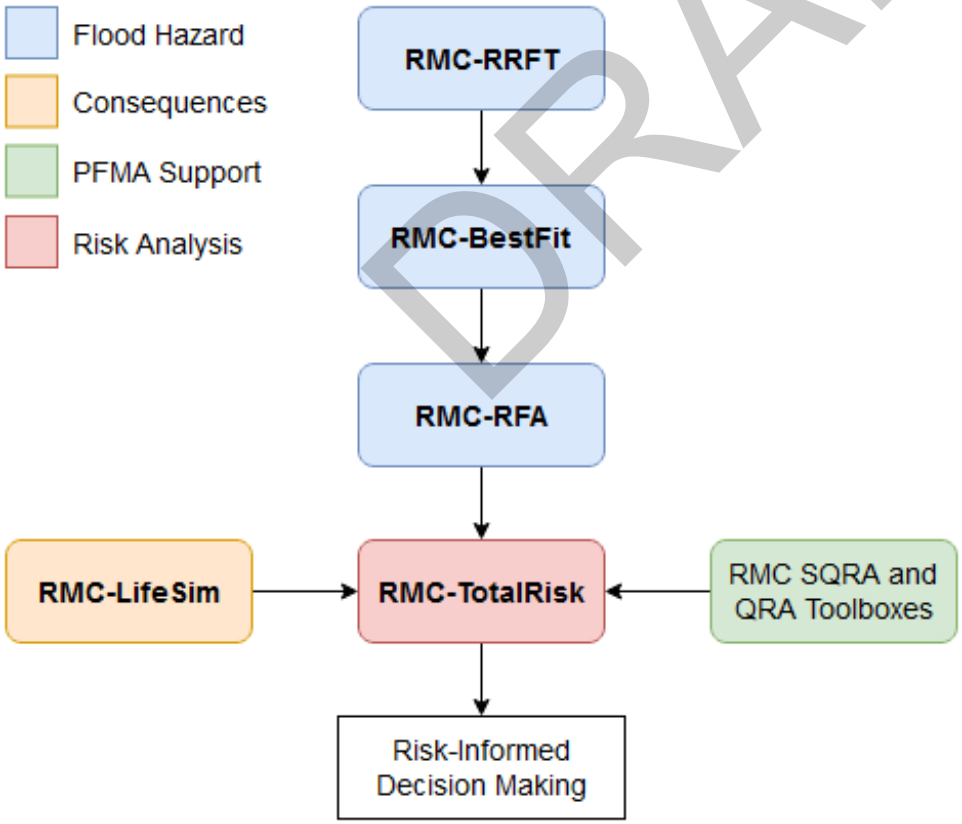


Figure 1 - Schematic of the RMC risk analysis software suite.

Flood hazard information can be estimated with the stochastic rainfall-runoff frequency tool (RRFT), the Bayesian estimation and fitting software (BestFit), and/or the reservoir frequency analysis software (RFA), and then imported to TotalRisk. These flood hazard tools are designed to work together or independently. For example, results from RRFT can be incorporated into BestFit, or entered directly into TotalRisk. Various semi-quantitative (SQRA) or quantitative risk assessment (QRA) toolboxes can be used to support the estimation of system response probabilities. Consequences can be estimated with and imported from LifeSim. RMC-TotalRisk then combines the hazard, system response, and consequences to calculate the system risk. TotalRisk can perform risk analysis for a single component, such as a dam or levee, or a complex system with multiple components, where each component can have multiple failure modes.

Overview of the Risk Analysis Framework

Risk has various definitions and interpretation among different industries, but it is generally understood to describe the probability and severity of an adverse event [2]. Risk can also be defined in terms of an expected value and deviations from the expected value. For example, in the financial industry, the standard deviation is a popular risk measure for stock returns [3]. Flood risk management investment decisions are typically made from a *risk neutral* perspective based on average annual net benefits [4]. As such, flood risk can be formally defined as the expected value of consequences $\mathbb{E}[C]$, which is calculated as:

$$\mathbb{E}[C] = \int_{-\infty}^{\infty} C(x) \cdot f\big(C(x)\big) \cdot dx \qquad \textit{Equation 1}$$

where x is the hazard level (e.g., flood discharge or water level); $C(x)$ determines the consequences, such as property damage or life loss, for the hazard level x; and $f\big(C(x)\big)$ is the probability density function (PDF) of the consequences occurring. The probability distribution of consequences can be defined as a function of hazard, system response, and consequence functions:

$$\mathbb{E}[C] = \int_{-\infty}^{\infty} f_x(x) \cdot F_R(x) \cdot C_R(x) \cdot dx \qquad \textit{Equation 2}$$

where $f_x(x)$ is the probability density function of the hazard (e.g., annual maximum water level or ground acceleration), $F_R(x)$ is the system response function (i.e., probability of failure given hazard level x), and $C_R(x)$ is the consequences given the hazard level x and the system response.

In practice, the risk integral for a dam or levee is often calculated numerically by annualizing and discretizing the hazard. The risk of failure using discrete hazard levels follows from Equation 2 and is defined as:

$$\mathbb{E}[C_F] = \sum_i P(x_i) \cdot P(F|x_i) \cdot C_F(x_i) \qquad \textit{Equation 3}$$

where $P(x_i)$ is the probability of the hazard level x_i; $P(F|x_i)$ is the conditional probability of failure given the hazard level x_i; and $C_F(x_i)$ is the consequence of failure given the hazard level x_i. Equation 3 is often written semantically to convey the risk equation as:

$$Risk\ of\ Failure\ = P(Hazard) \times P(Failure|Hazard) \times Consequences\ of\ Failure \qquad \textit{Equation 4}$$

where the risk of failure is equal to the probability of the hazard level, $P(Hazard)$, multiplied by the probability of failure given the hazard level, $P(Failure|Hazard)$, multiplied by the consequences of failure at the hazard level, $Consequences\ of\ Failure$.

In the risk analysis of dams and levees, the hazard is typically annualized by the annual maximum peak water surface elevation (WSE) or annual maximum peak ground acceleration (PGA) [5]. Annualizing the hazard results in an

annualized estimate for the expected consequences (e.g., average annual life loss or expected annual damage). Annualizing the risk estimate is not required in TotalRisk.

The system response function is conditioned on the hazard. Consequences are conditioned on both the hazard and the system response. Other parameters such as discharge, duration, velocity, magnitude, and ground motion time history can also be important for certain failure modes, such as erosion or liquefaction. Similarly, consequences may depend on breach location, size, and discharge along with warning and evacuation. When formulating a risk model, the risk analyst should select the most appropriate hazard variable and consider any additional effects and contributions that other relevant characteristics of the hazard may have on the system response and consequence functions.

The conceptual risk analysis process for a levee with a single failure mode and a single system component is shown in Figure 2 below. Beginning in the top left of the figure, the flood hazard is defined by an annual maximum peak flow-frequency distribution that is estimated using flood-frequency analysis methods. Next, moving to the top right, peak flow is then transformed to a WSE using a stage-discharge rating curve, which is estimated using a hydraulic model. Then, moving to the bottom right, the system response function is defined by a probability of failure given WSE, often derived from engineering analysis and expert elicitation methods. And finally, moving to the bottom left, the consequences given failure are estimated as a function of WSE. The expected annual consequences are computed by numerically integrating over these functions following Figure 2. Greater details on the mathematics of risk analysis are provided in the **Quantitative Risk Analysis** chapter. Additional details on risk analysis for flood risk management can be found in [6], [5], and [7].

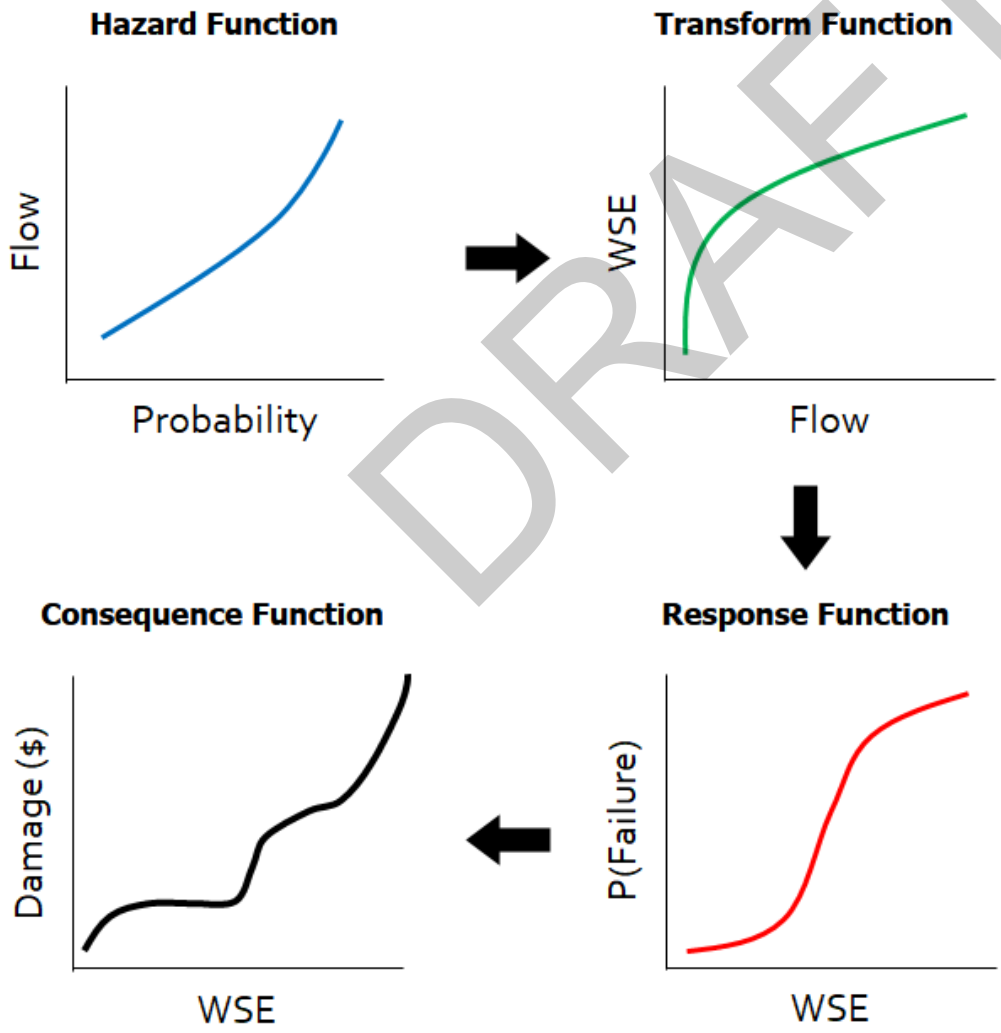


Figure 2 - Levee risk analysis process for a single failure mode and a single system component.

Risk Analysis Inputs

Figure 2 illustrates the key inputs for a single failure mode for a single system component. Starting from the top left of Figure 2 and moving clockwise, the key inputs are as follows:

- **Hazard Function**: A hazard function describes the exceedance probabilities of various hazard levels. Hazard functions are also commonly referred to as *frequency curves*. Examples include annual maximum peak flow-frequency, peak reservoir pool stage-frequency, and peak ground acceleration.

- **Transform Function**: A transform function can be used to transform (or convert) the hazard levels from one type of function to another. For example, a peak flow-frequency function can be transformed to a stage-frequency function using a flow-to-stage rating curve. Transform functions are optional inputs in TotalRisk.

- **System Response Function**: A system response function describes the conditional probability of failure for various hazard levels, such as water surface elevations. System response functions are sometimes referred to as *fragility curves*. The system response function defines the failure mode in RMC-TotalRisk.

- **Consequence Function**: A consequence function describes the consequences of failure or non-failure for various hazard levels, such as annual maximum peak water surface elevations. Consequence functions are also sometimes referred to as *damage functions*.

Natural Variability and Knowledge Uncertainty

There are two primary components of randomness that can be modeled in a quantitative risk analysis: natural variability and knowledge uncertainty. Natural variability is best described as the effect of randomness and is a function of the system [8]. This is also referred to as *aleatory uncertainty*. It is not reducible through either study or further measurement. For example, a peak flow-frequency curve describes the natural variability in peak flow. Another example is the distribution of consequences for various hazard levels, such as water surface elevation.

Knowledge uncertainty is the lack of knowledge about the parameters or processes that characterize the system being modeled. This is also referred to as *epistemic uncertainty*. Knowledge uncertainty can be reduced through further measurement or study. For example, the confidence intervals, or uncertainty bounds, around a peak flow-frequency curve describe the knowledge uncertainty in the statistical parameters of the peak flow-frequency curve. Another example is the uncertainty bounds around the distribution of life loss consequences, which describe the knowledge uncertainty in modeled warning times and evacuation rates in the flood inundated area.

There are two primary sources of knowledge uncertainty in quantitative risk analysis: sampling uncertainty and model uncertainty. First, the observed sample of large flood and seismic events, the structural performance during those events, and the associated flood damages is usually small. The relatively small sample sizes result in knowledge uncertainty pertaining to the true input probability distributions and corresponding consequence exceedance probabilities. This sampling uncertainty is a property of the effective sample size of the random variable; as the sample size increases, the knowledge uncertainty decreases.

Secondly, a probability distribution or simulation model does not always fully explain the variability of a particular input variable. This is referred to as model uncertainty, and it is due to the inherent assumptions in the formulation of the mathematical model itself, and our inability to identify the best fitting model.

In RMC-TotalRisk, the user-defined input functions model the natural variability of the system being evaluated. The input functions can be defined with either parametric or nonparametric methods, with or without uncertainty. Uncertainty about an input function defines the knowledge uncertainty. For the remainder of this document, knowledge uncertainty will be referred to as just *uncertainty*.

Risk Analysis

A risk analysis in RMC-TotalRisk is defined through a diagram as shown in Figure 3 below. The diagram provides an intuitive way to create and connect the various components of the modelled system. Figure 3 shows a single system component for a dam safety risk analysis. There is a non-failure mode, shown at the top of the diagram with the purple line, that connects the hazard function to the non-failure consequences, without any system response. For typical dams and levees, there will often be consequences even if the structure does not fail. For example, during a major flood event, a dam could activate the emergency spillway, or a levee could overtop, preventing them from reducing flood damage. The non-failure mode is used to model the risk of non-failure. There are two failure modes: 1) An internal erosion failure mode, labelled PFM 1, shown in the center of the diagram connects the hazard at Dam A to

the PFM 1 response function and the PFM 1 consequences; and 2) An overtopping failure mode, labelled PFM 2, shown in the bottom of the diagram with the same respective connections.

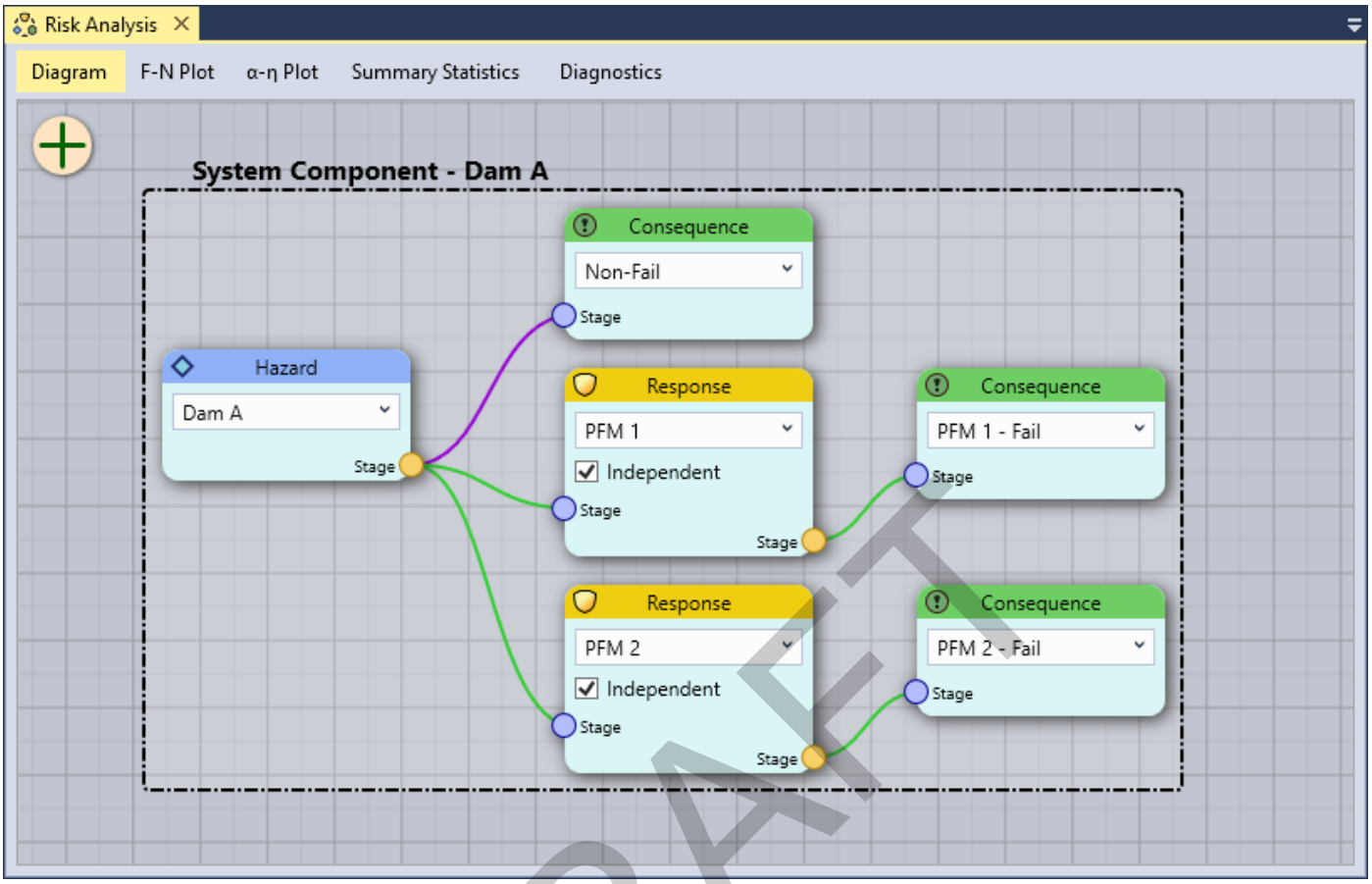


Figure 3 - RMC-TotalRisk risk diagram.

The system components are identified and labelled by the selected hazard function. The failure modes within a component are identified and labelled by the selected response functions. RMC-TotalRisk permits an unlimited number of failure modes per component. However, a single system is limited to 20 components due to virtual memory and computer runtime limitations. For example, the system risk of a watershed comprising up to 20 dams, each with dozens of failure modes, can be assessed.

Dependency between failure modes and system components can be defined in TotalRisk. There is rarely enough data to estimate the dependency between failure modes. Instead, failure modes can be modelled as perfectly independent, or perfectly negatively dependent. Perfect independence is an upper bound when failure modes are positively correlated. Whereas perfect negative dependency is an upper bound when failure modes are negatively correlated [9]. Likewise, dependency between system components can be set as perfectly independent, positive, or negatively dependent. There is also an option to set the dependency between system components with a user defined correlation matrix.

After the inputs and dependency options have been selected, the risk for each individual failure mode, each system component, and the overall system can be computed. The overall risk and Monte Carlo simulation framework employed by RMC-TotalRisk is illustrated in Figure 4 below.

The following chapters provide technical details on the risk analysis inputs, and the available options for each. The **Quantitative Risk Analysis** chapter provides complete details on how the inputs are used to estimate risk in RMC-TotalRisk.

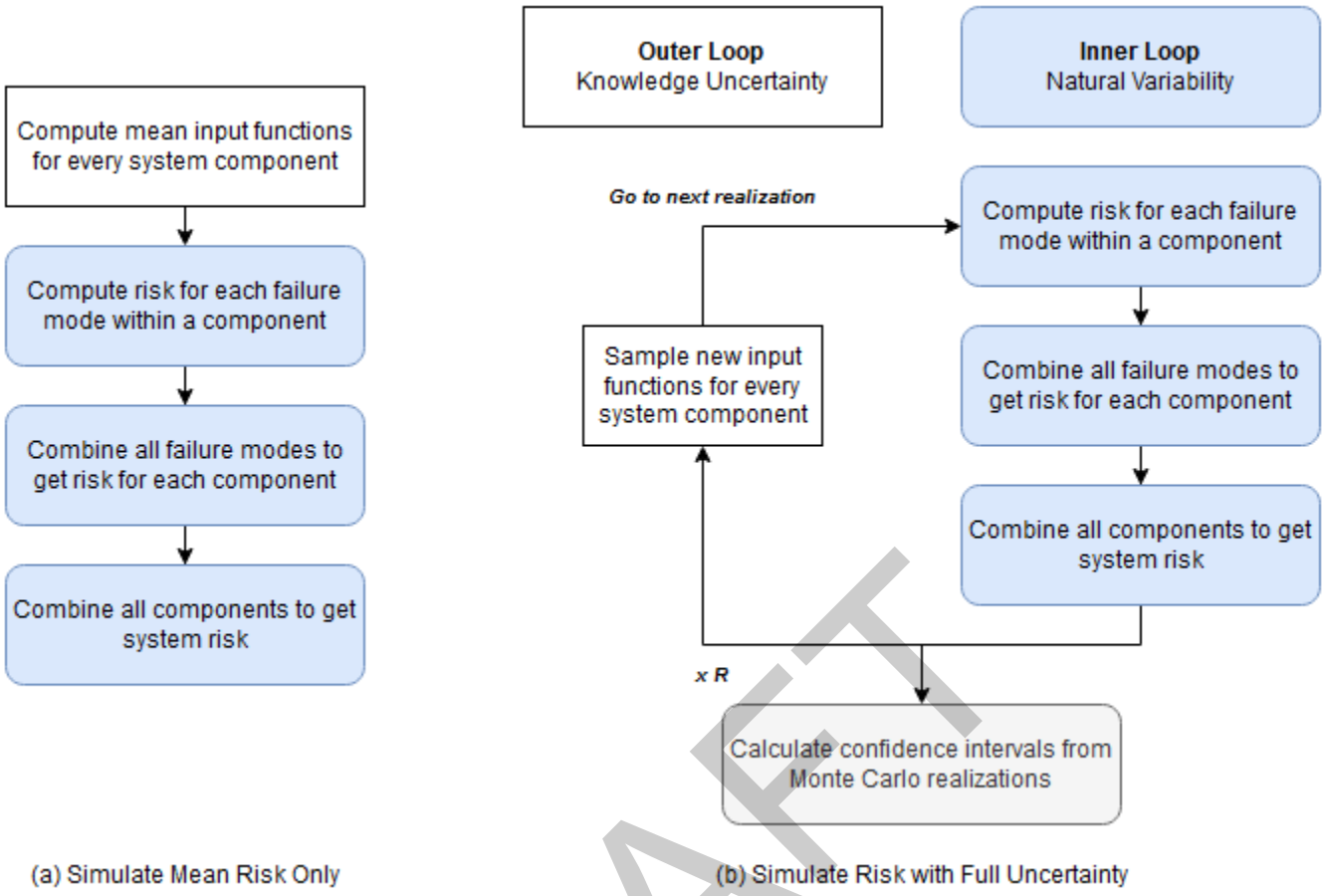


Figure 4 - Flowchart of the RMC-TotalRisk simulation options: (a) Simulate mean risk only, and (b) Simulate risk with full uncertainty.

Applicability

RMC-TotalRisk was developed to support risk assessments within the USACE Flood Risk Management, Planning, and Dam and Levee Safety Communities of Practice. The risk analysis capabilities and features were developed in accordance with the following USACE risk management guidance:

- Engineer Regulation (ER) 1105-2-100, Planning Guidance Notebook [10]
- ER 1105-2-101, Risk Assessment for Flood Risk Management Studies [2]
- ER 1110-2-1156, Safety of Dams – Policy and Procedures [11]
- Engineer Manual (EM) 1110-2-1619, Risk-based Analysis for Flood Damage Reduction Studies [7]
- Best Practices in Dam and Levee Safety Risk Analysis [5]

RMC-TotalRisk is a powerful and general-purpose risk analysis software, capable of estimating risk for a variety of systems. Typical applications in science and engineering include flood risk management, but TotalRisk can be used for any risk analysis application that has a hazard and a consequence.

Hazard Functions

In RMC-TotalRisk, a hazard function is defined by the exceedance probabilities of various hazard levels, such as annual maximum peak flow or water surface elevation. Hazard functions are also commonly referred to as *frequency curves*. In the risk assessment of dams and levees, the annual maximum peak water surface elevation (or stage) or the annual maximum peak ground acceleration are typically the primary hazard parameters for evaluating a potential failure mode [5] [12]. As such, the hazard functions will commonly describe the annual exceedance probability (AEP) of these hazard levels.

Hazard functions can be defined as either a parametric or nonparametric distribution. A parametric distribution has a theoretical mathematical form, with input parameters, that describes how the probabilities are distributed over the values of a random variable. Parametric distributions are sometimes referred to as *theoretical* or *analytical* distributions [7] [13]. A nonparametric probability distribution makes very few assumptions about the underlying theoretical model, and it is based on empirical or simulated data. Nonparametric distributions are sometimes referred to as *nonanalytical*, *empirical*, *graphical*, or *tabular* distributions [7] [13] [14].

All hazard functions in RMC-TotalRisk are modeled as a continuous random variable (see **Appendix A**). As such, each have a probability density function (PDF), $f(x)$, a cumulative distribution function (CDF), $F(x)$, a survival function, $S(x) = 1 - F(x)$, and an inverse CDF, $F^{-1}(p)$. All hazard functions are numerically integrated between a probability of 10^{-16} and $1 - 10^{-16}$, which is sufficiently close to 0 and 1 (see **Appendix D**). This is done to make the risk results collectively exhaustive.

The following subsections describe the various hazard function options in RMC-TotalRisk.

Parametric

A hazard function can be defined with a parametric probability distribution. The following parametric distribution options are available for parametric hazard functions:

- Exponential
- Gamma
- Generalized Extreme Value
- Generalized Logistic
- Generalized Normal
- Generalized Pareto
- Gumbel (EV1)
- Kappa-4
- Log-Normal
- Log-Pearson Type III
- Pearson Type III

For details on these distributions, please see **Appendix E**. Parametric distributions can be estimated from distribution fitting software, such as *RMC-BestFit*, *HEC-SSP*, or other various commercial software.

TotalRisk uses the parametric bootstrap [15] [16] to quantify uncertainty in a parametric hazard function. The bootstrap procedure involves the following general steps:

1. Randomly sample n hazard levels from the user-defined probability distribution, or *parent distribution*, where n is equal to the effective record length (ERL). This is called the *bootstrap sample*.

2. Estimate a new distribution from the bootstrap sample. The distribution can be estimated with product moments, linear moments, or maximum likelihood. See [17], [18], and [19] for more details on these estimation and fitting methods for distributions.

3. Record quantiles for desired nonexceedance probabilities, and any other output of interest.

4. Repeat steps 1 through 3 for a sufficient large number of realizations, R. Then, derive confidence intervals by computing percentiles from the bootstrapped array for the desired output.

The parametric bootstrap is more formally described in Algorithm 1 below. The user must enter the ERL, which is a measure of information content in the fit of the distribution. The longer the ERL, the less uncertainty and narrower the confidence intervals. The ERL is sometimes referred to as the *equivalent record length* or *effective sample size*. Many definitions exist, but in the context of flow-frequency, ERL is typically defined as the number of exact data needed to produce the same confidence intervals given a combination of exact and censored data [20] .Guidance on selecting an appropriate ERL is provided in [7] and [20].

Algorithm 1 – Parametric Bootstrap Analysis

$R \leftarrow$ number of bootstrap realizations
$n_{erl} \leftarrow$ effective record length (ERL)
$n_p \leftarrow$ number of user-specified exceedance probabilities
for $i \leftarrow 1$ to R **do**
  Sample at random with replacement n_{erl} hazard levels $\boldsymbol{x_i^*} \leftarrow \left(x_1^*, \cdots, x_{n_{erl}}^*\right)$ from parent distribution $f(\theta)$.
  Estimate and record the distribution parameters θ_i^* given the bootstrap sample $\boldsymbol{x_i^*}$.
  for $j \leftarrow 1$ to n_p **do**
    Record quantile for specified exceedance probability $q_{ij}^* = F^{-1}\left(1 - p_j | \theta_i^*\right)$.
  end for
end for
Estimate confidence intervals and mean hazard curve from all bootstrap realizations $\{\boldsymbol{q_1^*}, \cdots, \boldsymbol{q_R^*}\}$, for all exceedance probabilities.

An example of the modeled parametric bootstrap uncertainty for a parametric hazard is provided in Figure 5. This figure shows a Log-Pearson Type III distribution for peak flows (cfs) with a mean (of log) of 3.0, a standard deviation (of log) of 0.2, a skew (of log) of 0.5, an ERL of 100, and 10,000 bootstrap realizations.

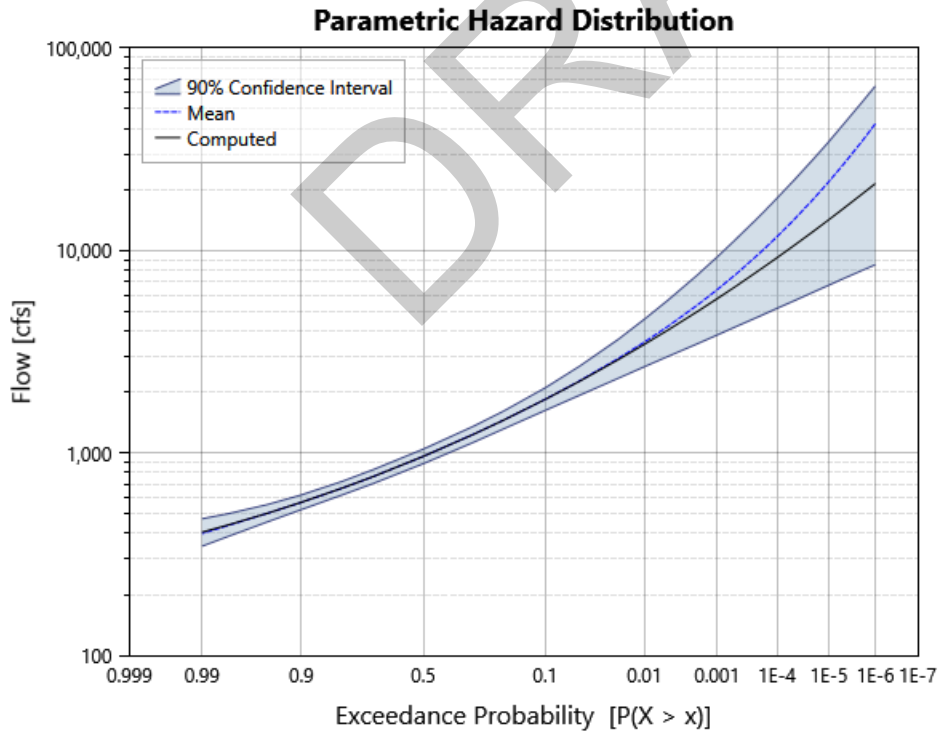


Figure 5 - Example of a parametric hazard function.

By default, the bootstrap analysis has 10,000 realizations because this is the maximum number of Monte Carlo realizations allowed in the risk analysis. When estimating risk with full uncertainty, one of the 10,000 bootstrapped parameter sets is selected at random each Monte Carlo realization (see Algorithm 2). Then, a new parametric probability distribution is created from the parameter set. Hazard levels and exceedance probabilities can then be computed within the risk analysis for that realization.

Algorithm 2 – Parametric Function Uncertainty Analysis

$R \leftarrow$ number of Monte Carlo realizations
$R_\theta \leftarrow$ number of parameter sets
Where $R \leq R_\theta$
for $i \leftarrow 1$ to R **do**
  if sub function within a composite function **then**
    Sample uniformly at random with replacement a percentile $r_i \sim \mathrm{U}(0,1)$.
    Sample a parameter set θ_i^* from list at index $1 + (floor)[r_i \cdot R_\theta]$
  else
    Sample a parameter set θ_i^* from list at index i
  end if
  Create new parametric hazard function H_i^* from sampled parameter set θ_i^*.
end for
Estimate confidence intervals and mean hazard curve from all sampled hazard functions $\{H_1^*, \cdots, H_R^*\}$.

RMC-BestFit

A parametric hazard function can be imported from the Bayesian estimation and fitting software, *RMC-BestFit*. Technical details on the *BestFit* software are provided in [21] [22] [23].

RMC-BestFit performs Bayesian estimation from a choice of thirteen parametric probability distributions. *RMC-BestFit* outputs the posterior mode and posterior predictive distributions and credible intervals. In addition, by default, *RMC-BestFit* outputs 10,000 posterior parameter sets which are used to propagate uncertainty in RMC-TotalRisk. Once imported, the *BestFit* results are stored in the RMC-TotalRisk study file. If the original *BestFit* results are updated, they will have to be re-imported.

When estimating risk with full uncertainty, one of the 10,000 posterior parameter sets is selected at random each Monte Carlo realization as described in Algorithm 2. Then, a new parametric probability distribution is created from the parameter set. Hazard levels and exceedance probabilities can then be computed within the risk analysis for that realization.

Nonparametric

Nonparametric hazard functions are defined with a tabular relationship of hazard levels and exceedance probabilities. The nonparametric survival function $S(x) = 1 - F(x)$ and inverse survival function $S^{-1}(p)$ are both computed using linear interpolation as follows:

$$S(x) = p_i + (p_{i+1} - p_i)\left(\frac{x - x_i}{x_{i+1} - x_i}\right) \qquad \textit{Equation 5}$$

$$S^{-1}(p) = x_i + (x_{i+1} - x_i)\left(\frac{p - p_i}{p_{i+1} - p_i}\right) \qquad \textit{Equation 6}$$

where $S(x)$ is the survival function of the variable X; $S^{-1}(p)$ is the inverse survival function where p is an exceedance probability; and there is an array of continuous hazard values $\boldsymbol{x} = \{x_1, x_2, \ldots, x_n\}$ for $x_i < x < x_{i+1}$with exceedance probabilities $\boldsymbol{p} = \{p_1, p_2, \ldots, p_n\}$ with $0 \leq p_i \leq 1$ and $p_i > p > p_{i+1}$.

In general, extrapolation will not be performed for p values outside the user-defined range of hazard levels. Instead x values will be flatlined on each side of the table. If $p > p_1$, then x_1 will be returned. Likewise, if $p < p_n$, then x_n will be returned. However, there is an option for extrapolation to a rare exceedance probability provided in the HEC-FDA nonparametric option discussed in the next section. All hazard functions are numerically integrated between a probability of 10^{-16} and $1 - 10^{-16}$, which is sufficiently close to 0 and 1 (see **Appendix D**). Therefore, to improve the accuracy of the risk results and to avoid significant flatlining, it is important to input the tabular relationship so that there is sufficient coverage across probabilities.

There is often a benefit to applying a transform to the hazard and/or probability values to improve the accuracy of the linear interpolation. A logarithmic transform can be applied to the hazard. A logarithmic or Normal z transform can be applied to the exceedance probability.

For example, if the hazard levels increase exponentially in real-space, then they will increase linearly in log-space. In this case, a log-transform will improve the accuracy of the linear interpolation of hazard levels. When the hazard levels are log-transformed, the inverse survival function becomes:

$$S^{-1}(p) = \log x_i + (\log x_{i+1} - \log x_{i+1}) \left(\frac{p - p_i}{p_{i+1} - p_i} \right) \qquad \textit{Equation 7}$$

When the exceedance probabilities are Normal z transformed, the inverse survival function becomes:

$$S^{-1}(p) = x_i + (x_{i+1} - x_i) \left(\frac{\Phi^{-1}(p) - \Phi^{-1}(p_i)}{\Phi^{-1}(p_{i+1}) - \Phi^{-1}(p_i)} \right) \qquad \textit{Equation 8}$$

where $\Phi^{-1}(\cdot)$ is the inverse CDF of the standard Normal distribution.

Nonparametric – HEC-FDA

A nonparametric hazard function can be defined in the same way as the "less simple method" in the flood damage reduction analysis software, HEC-FDA [13]. In RMC-TotalRisk, this hazard function is just called the *Nonparametric* hazard function.

In this method, the user enters a tabular relationship of hazard levels and exceedance probabilities. The uncertainty in the hazard level for a given exceedance probability is derived using the asymptotic approximation for quantile variance. A detailed proof is provided in [14] and [24]. Additional details related to the HEC-FDA implementation are provided in [13]. The variance of a hazard level (quantile) for a given exceedance probability is estimated from:

$$\sigma_{x_i}^2 = \frac{(1 - p_i) \cdot p_i}{n \cdot f(x_i)^2} \qquad \textit{Equation 9}$$

where $\sigma_{x_i}^2$ is the variance of the hazard level x_i; p_i is the exceedance probability; n is the effective record length (ERL); and $f(\cdot)$ is the nonparametric PDF. As shown in **Appendix A**, the PDF $f(\cdot)$ is the derivative of the CDF:

$$f(x_i) = \frac{dF(x_i)}{dx_i} = \frac{f(x_i + h) - f(x_i - h)}{2h} \qquad \textit{Equation 10}$$

The derivative is evaluated using numerical differentiation with the two-point formula where h represents a small change in x_i. To avoid unreasonable PDF and variance results, the quantile variance $\sigma_{x_i}^2$ computed for the 0.01 exceedance probability is used for all quantiles larger ($p < 0.01$), and the $\sigma_{x_i}^2$ for the 0.99 exceedance probability is used for all quantiles smaller ($p < 0.99$).

Like the parametric hazard function, the user must enter the ERL, which is a measure of information content in the fit of the distribution. The longer the ERL, the less uncertainty and narrower the confidence intervals. The user can also choose to extrapolate to a rarer exceedance probability. For consistency with HEC-FDA, the default is to perform linear extrapolation out to $p = 0.0001$. If the user defined data already extends beyond $p = 0.0001$, no extrapolation will be performed. Also, by default, linear extrapolation back to $p = 0.999$ will be performed automatically if the user defined data does not already cover this exceedance probability.

The uncertainty in the hazard level is automatically set to be Ln-Normal distributed with a real-space mean, $\mu = x_i$, and standard deviation, $\sigma = \sigma_{x_i}$. This is done to be backwards compatible with existing HEC-FDA models that utilize a similar uncertainty routine. However, this type of nonparametric hazard function be too restrictive for complex scenarios, such as stage-frequency curves for dams. The *Tabular* hazard function described in the next section provides more flexibility for defining uncertainty.

The uncertainty analysis for all nonparametric hazard functions is performed as described in Algorithm 3 below. This routine describes the case where the hazard level has uncertainty. However, in the *Tabular* hazard function described in the next section, the user can instead choose to model the exceedance probabilities as uncertain while holding the hazard levels as fixed. In this case, the uncertainty analysis is performed in the same manner, but with random exceedance probabilities rather than random hazard levels. A new nonparametric hazard function is created each Monte Carlo realization, which is then used to derive hazard levels and exceedance probabilities in the risk analysis.

Algorithm 3 – Nonparametric (Tabular) Function Uncertainty Analysis

$R \leftarrow$ number of Monte Carlo realizations
$n \leftarrow$ number of user-specified hazard levels
for $i \leftarrow 1$ to R **do**
  Sample uniformly at random with replacement a percentile $r_i \sim \mathrm{U}(0,1)$.
  for $j \leftarrow 1$ to n **do**
    Generate a hazard level $x_j^* = F^{-1}(r_i|\theta_j)$ for tabular ordinate j.
  end for
  Create new tabular hazard function H_i^* from output $\boldsymbol{p} = \{p_1, p, \dots, p_n\}$ and $\boldsymbol{x}^* = \{x_1^*, x_2^*, \dots, x_n^*\}$.
end for
Estimate confidence intervals and mean hazard curve from all sampled hazard functions $\{H_1^*, \cdots, H_R^*\}$.

This uncertainty analysis approach is the same as the approach taken in HEC-FDA for *graphical* or *non-analytic* relationships. Complete details are provided in section G.3.3 in the HEC-FDA user's manual [13]. This algorithm is restrictive in terms of the possible shapes of the nonparametric distribution that can be randomly generated. This could lead to a slight overestimation or underestimation in the variance of the risk results. However, as discussed in [13], generalizing the shape of the distribution requires a parametric representation. In the absence of a parametric component, this is currently the best algorithm available for nonparametric uncertainty analysis.

Tabular

As with the previous nonparametric option, for this method, the user enters a tabular relationship of hazard levels and exceedance probabilities. Extrapolation will not be performed for values outside the user-defined range of hazard levels.

In many cases, tabular hazard functions will be derived from external simulation software, such *RMC-RRFT* or *RMC-RFA*. Or for seismic hazards, nonparametric functions are often derived from a probabilistic seismic hazard analysis (PSHA) [5]. An example of a seismic hazard function is shown in Figure 6 below. In this case, only the mean hazard distribution has been entered and there is no uncertainty. Figure 7 below shows an example of a reservoir stage-frequency curve from *RMC-RFA* entered with uncertainty in the exceedance probabilities.

The user can choose to model the exceedance probabilities as uncertain while holding the hazard levels as fixed, or vice versa. A distribution must be selected to define uncertainty. The parameters for the selected distribution must be entered for every ordinate in the tabular data. This allows the parameters to vary with hazard levels. The uncertainty at each hazard level (or exceedance probability) must be entered such that the confidence intervals are monotonically increasing with increasing hazard levels.

The following distribution options are available for all tabular functions in RMC-TotalRisk:

- Deterministic (no uncertainty)
- Generalized Beta
- Ln-Normal
- Normal
- PERT
- PERT-Percentile
- PERT-PercentileZ (when probability is uncertainty)
- Triangular
- Truncated Normal
- Uniform

For details on these distributions, please see **Appendix E**.

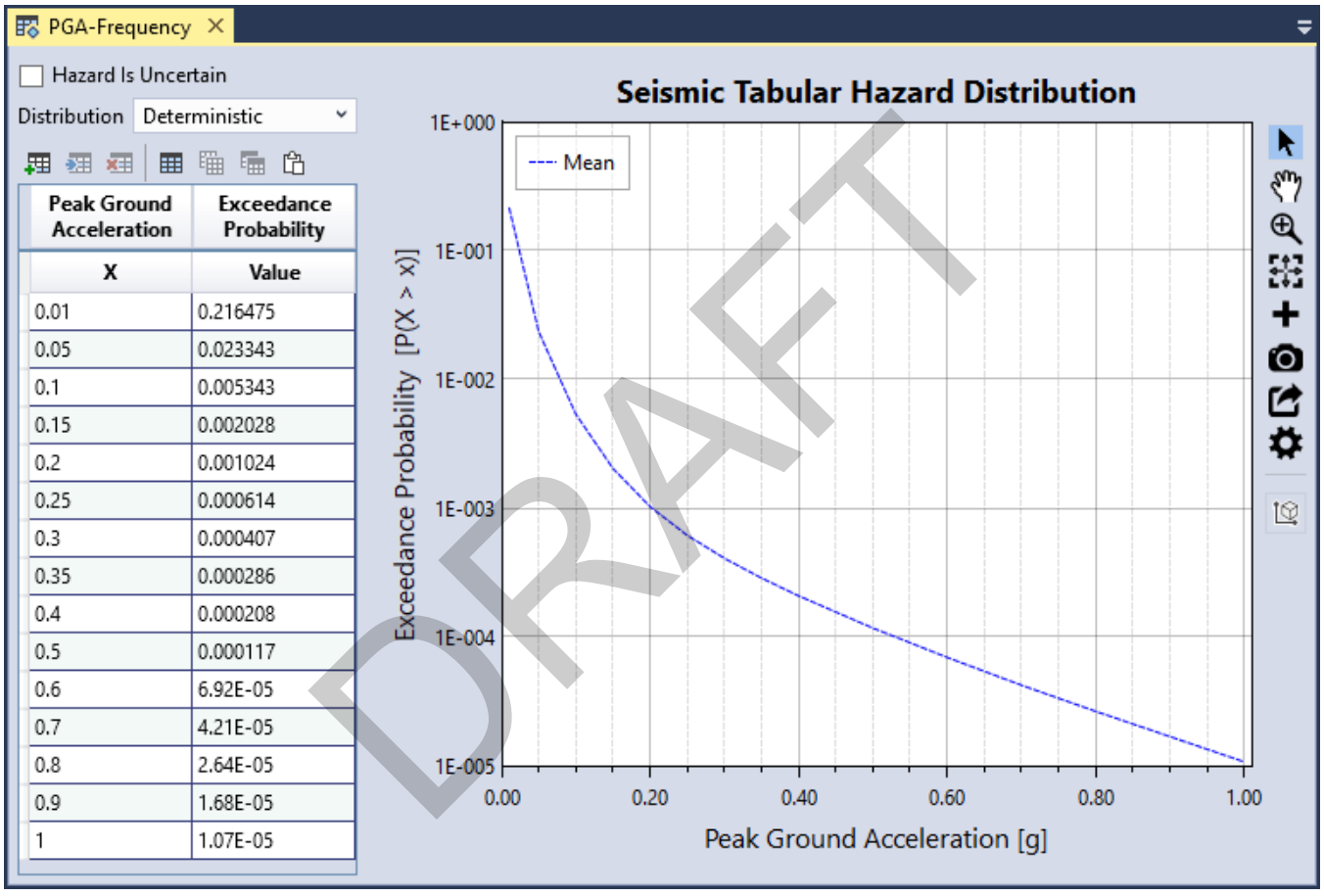


Figure 6 - Example of a seismic tabular hazard function.

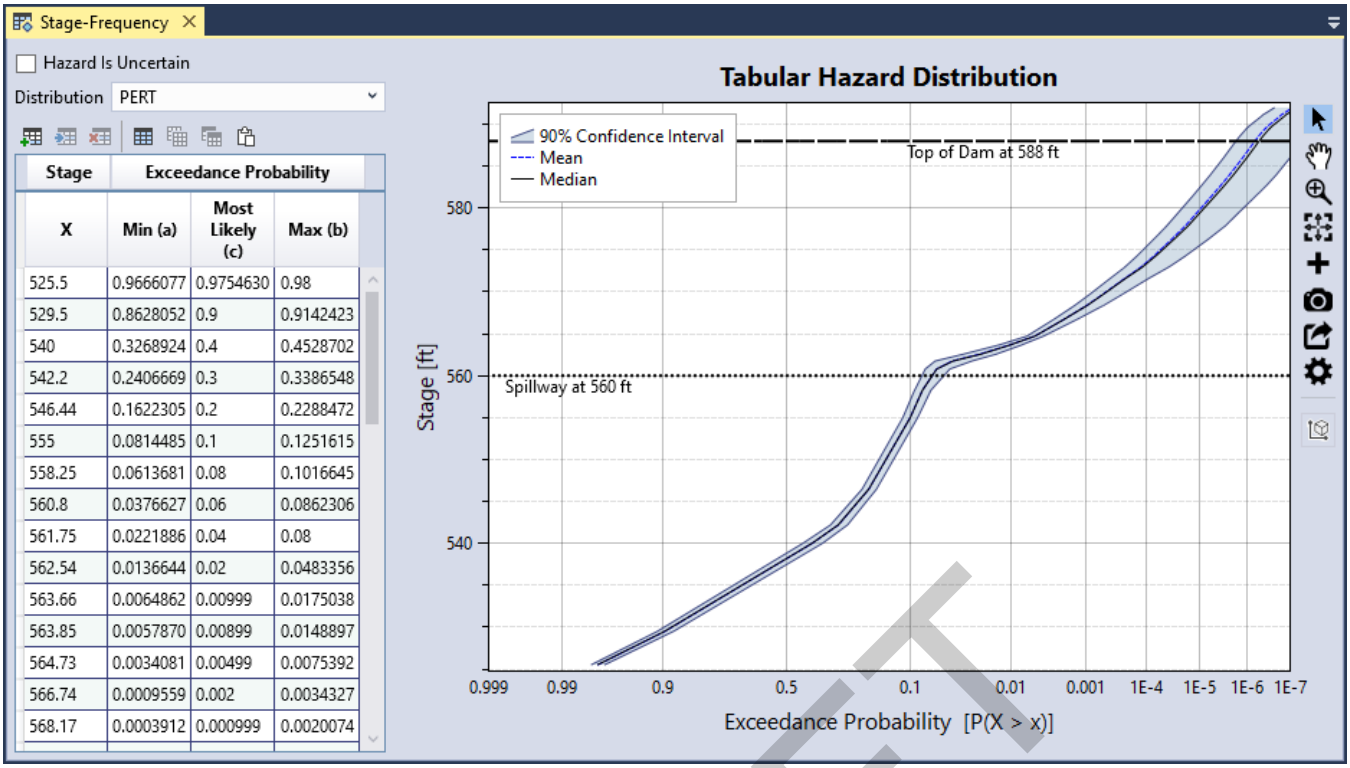


Figure 7 - Example of a stage-frequency tabular hazard function with uncertainty.

RMC-RFA

A hazard function can be imported from the reservoir frequency analysis software, *RMC-RFA*. Technical details on the *RMC-RFA* software can be found in [12], [25], and [26]. In *RMC-RFA* version 1.0.1, there are three simulation types:

- Full uncertainty
- Expected curve only
- Median curve only

And there are three output frequency curve options:

- Stage-frequency curve
- Peak discharge-frequency curve
- Discharge-duration-frequency curve

In RMC-TotalRisk, the expected curve and median frequency curve results are imported as a tabular hazard function with no uncertainty. The full uncertainty frequency curve results are imported as a tabular hazard function, where the uncertainty in the hazard levels for each exceedance probability is modeled with a PERT distribution based on the imported confidence interval values. For example, if a RFA simulation produced 90% confidence intervals, then the minimum of the PERT will be the 5th percentile and the maximum will be the 95th percentile. The mode is defined using the median frequency curve. The use of the PERT in this manner will reduce the confidence interval widths from *RMC-RFA*. Please see the Tabular hazard function section for more details on the tabular function uncertainty analysis. This approach with the PERT distribution was taken due to the challenges of importing results from *RFA* version 1.0.1. Importing stochastic simulation results will be more robust in future versions of *RFA* and TotalRisk. Once imported, the *RFA* results are stored in the RMC-TotalRisk study file. If the original *RFA* results are updated, they will have to be re-imported.

Composite

A composite hazard function can be created by assigning weights (or likelihoods) to a list of hazard functions as follows:

$$F(x) = \sum_{i=1}^{n} \omega_i \cdot F_i(x) \qquad \textit{Equation 11}$$

where $F_i(\cdot)$ is the CDF for hazard function i; and ω_i is the weight or likelihood of hazard function i, with $0 \leq \omega_i \leq 1$ and $\sum_{i=1}^{n} \omega_i = 1$. This type of composite function is traditionally referred to as a *mixture distribution* [27].

In dam safety, it is common practice to evaluate various gate failure or debris blockage scenarios as separate simulations in *RMC-RFA*, and then assigning a likelihood to each scenario. For example, the risk analysis team might evaluate the stage-frequency assuming 0%, 10%, 25%, and 50% reduction in spillway discharge capacity due to debris blockage. Then, the likelihood of each of these scenarios would be elicited by a group of experts.

Figure 8 shows an example of a composite stage-frequency curve for four debris blockage scenarios. In this example, as shown in Table 1, 0% blockage was given a weight of 0.25, 10% blockage a weight of 0.4, 25% blockage a weight of 0.2, and 50% blockage a weight of 0.15. A small amount of debris blockage is considered the most likely scenario during a large flood.

Table 1 - Example of debris blockage weights

% Blockage	% Range	Weight
0	0 to 5	0.25
10	5 to 12.5	0.40
25	12.5 to 37.5	0.20
50	>37.5	0.15

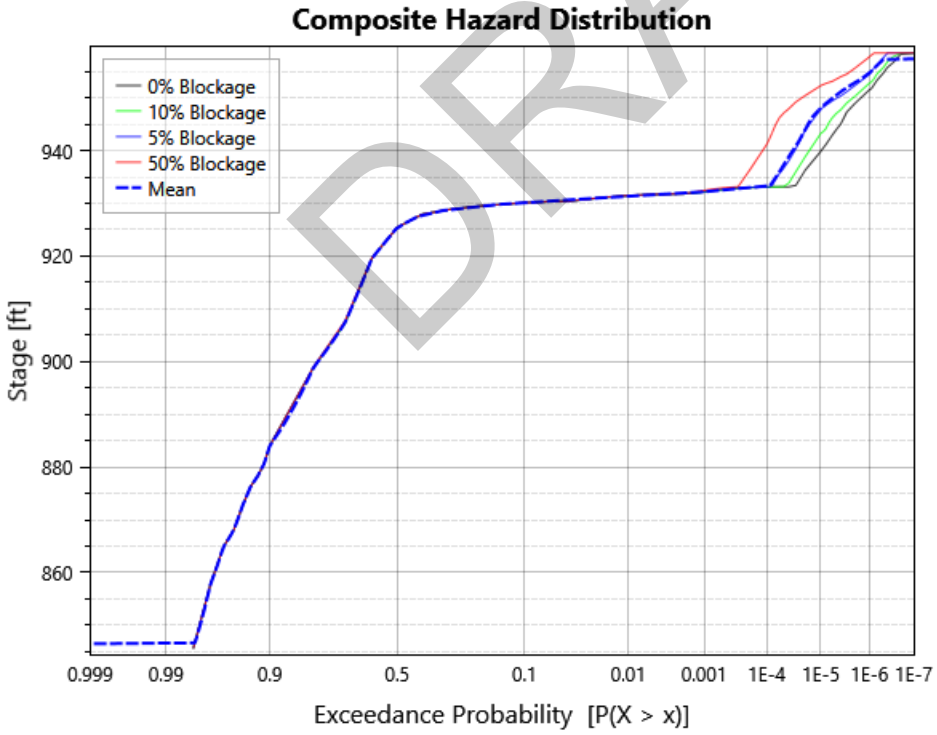


Figure 8 - Example of a composite stage-frequency hazard function for combining debris loading scenarios (weighted mixture).

Rather than using weights, a composite hazard function can alternatively be created by combining a list of hazard functions using the probability of union, assuming statistical independence between functions. It is common in hydrology that annual maximum flood events are composed of events that arise from distinctly different processes. For example, in California, annual maximum inflows to a dam might be driven by atmospheric river driven rainfall events in the winter or snowmelt events in the spring. Floods arising from rainfall or snowmelt will often have distinctly different probability distributions. In this situation, the annual maximum flood event should be viewed as the maximum of the annual maximum rainfall driven events R and the annual maximum of the snowmelt driven events S [28].

$$F(x) = P[X = \max(R, S) \leq x] = F_R(x) \cdot F_S(x) \qquad \textit{Equation 12}$$

where $F_R(\cdot)$ is the CDF of annual maximum rainfall driven floods; and $F_S(\cdot)$ is the CDF of annual maximum snowmelt driven floods. This assumes the rainfall and snowmelt driven events are statistically independent from one another. Equation 12 can be written more generally for multiple hazard functions as:

$$F(x) = \prod_{i=1}^{n} F_i(x) \qquad \textit{Equation 13}$$

where $F_i(\cdot)$ is the CDF for hazard function i. In USACE, Equation 12 and Equation 13, respectively, are typically written in terms of the probability of union of exceedance probabilities [29] as follows:

$$S(x) = P[X = \max(R, S) > x] = S_R(x) + S_S(x) - S_R(x) \cdot S_S(x) \qquad \textit{Equation 14}$$

$$S(x) = 1 - \prod_{i=1}^{n} [1 - S_i(x)] \qquad \textit{Equation 15}$$

where $S_R(\cdot)$ is the survival function, $1 - F_R(\cdot)$, of annual maximum rainfall driven floods; and $S_S(\cdot)$ is the survival function of annual maximum snowmelt driven floods; and $S_i(\cdot)$ is the survival function, $1 - F_i(\cdot)$, for hazard function i. Please see **Appendix A** for more details on the probability of union.

Figure 9 shows an example of a composite flow-frequency curve for combining rainfall and snowmelt driven flow-frequency curves. When combining hazard functions with a weighted mixture (Equation 11), the composite function will be the weighted average of the list of marginal hazard functions. However, when combining hazard functions with the probability of union (Equation 15), the composite function will always be greater than or equal to the most severe of the marginal hazard functions.

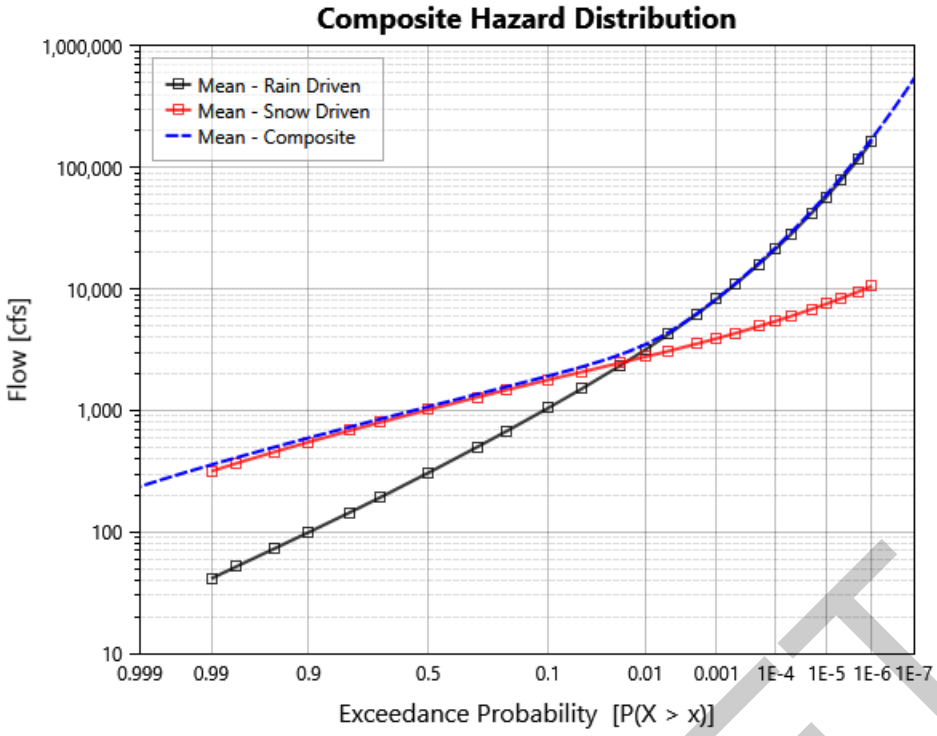


Figure 9 - Example of a composite flow-frequency hazard function for combining two flood driving mechanisms (probability of union).

The uncertainty analysis for the composite hazard functions is performed as described in Algorithm 4. This algorithm describes the routine given a single hazard level of interest, but it can easily be extended to include a list of hazard levels to evaluate. When simulating risk with full uncertainty, a new composite hazard function is created each Monte Carlo realization, which is then used to derive hazard levels and exceedance probabilities in the risk analysis. The confidence interval shown in Figure 10 below were created using this algorithm.

Algorithm 4 – Composite Hazard Function Uncertainty Analysis

$R \leftarrow$ number of Monte Carlo realizations
$n \leftarrow$ number of hazard functions
$x \leftarrow$ hazard level to evaluate
for $i \leftarrow 1$ to R **do**
  if mixture **then** $p_i^* = 0$ **else** $p_i^* = 1$
  for $j \leftarrow 1$ to n **do**
    Sample uniformly at random with replacement a percentile $r_i \sim \mathrm{U}(0,1)$.
    Sample a random hazard function F_j^* given r_i
    if mixture **then** $p_i^* \leftarrow p_i^* + \omega_j \cdot F_j^*(x|r_i)$ **else** $p_i^* \leftarrow p_i^* \cdot F_j^*(x|r_i)$
  end for
end for
Estimate confidence intervals and mean non-exceedance probability for x from $\{p_1^*, \cdots, p_R^*\}$.

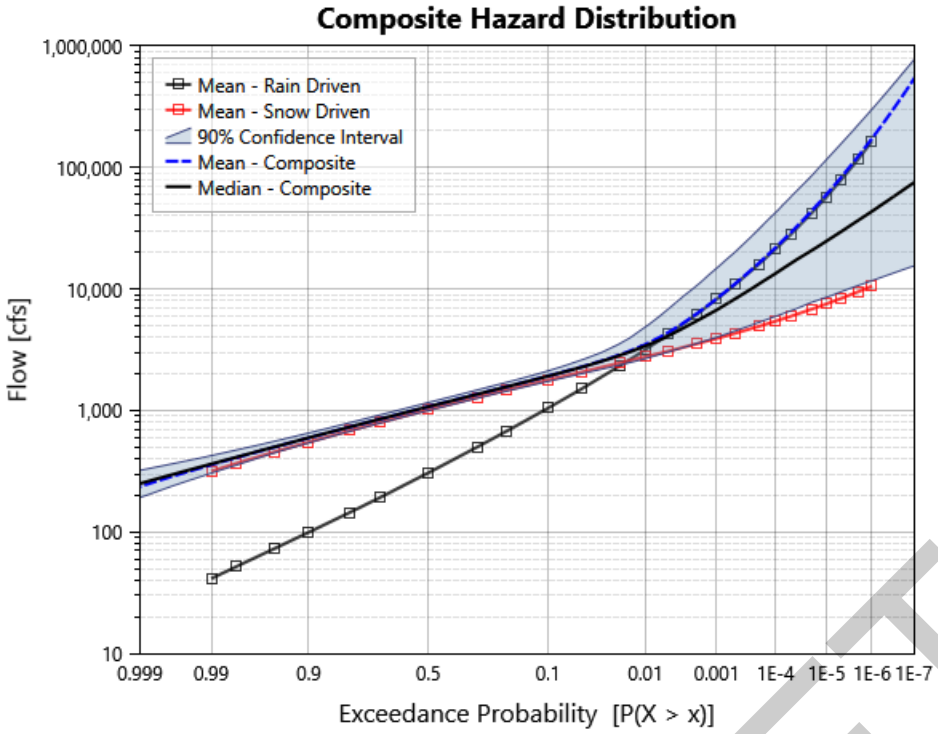


Figure 10 - Example of a composite flow-frequency hazard function for combining two flood driving mechanisms (probability of union) with uncertainty.

Transform Functions

In RMC-TotalRisk, a transform function can be used to transform (or convert) the hazard levels from one type of function to another. Transforming hazard levels is also referred to as *function composition*. In mathematics, function composition is an operation that takes two functions f and g, and creates a function h as follows:

$$h(x) = g\big(f(x)\big) \qquad \textit{Equation 16}$$

For example, a peak flow-frequency function $f(\cdot)$ can be transformed to a stage-frequency function $h(\cdot)$ using a flow-to-stage rating curve $g(\cdot)$. Transform functions can be used to transform hazard levels for hazard functions, other transform functions, and system response functions. The following subsections describe the various transform function options in RMC-TotalRisk.

Linear

Hazard levels can be transformed using a simple linear equation:

$$h(x) = y = \alpha + \beta x + \varepsilon \qquad \textit{Equation 17}$$

where x is the hazard level; y is the transformed hazard level; α is the intercept coefficient; β is the slope coefficient; and ε is the model error, or residual, which is assumed to be normal and independently distributed (NID) with zero mean and standard error σ, or $\varepsilon \sim \Phi(0, \sigma)$. When simulating risk with full uncertainty, linear transform functions are sampled from a Normal distribution based on the specified standard error as shown in Algorithm 5.

Algorithm 5 – Linear Transform Function Uncertainty Analysis

$R \leftarrow$ number of Monte Carlo realizations
$x \leftarrow$ hazard level to evaluate
for $i \leftarrow 1$ to R **do**
  Sample uniformly at random with replacement a percentile $r_i \sim \mathrm{U}(0,1)$
  $y_i^* \leftarrow \alpha + \beta x + \Phi^{-1}(r_i|0, \sigma)$
end for
Estimate confidence intervals and mean transformed hazard for x from $\{y_1^*, \cdots, y_R^*\}$.

The parameters of this linear equation can be estimated using the ordinary least squares (OLS) method for simple linear regression as follows:

$$\beta = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\,(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad \textit{Equation 18}$$

$$\alpha = \bar{y} - \beta x \qquad \textit{Equation 19}$$

$$\sigma = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2} \qquad \textit{Equation 20}$$

Where $\boldsymbol{x} = \{x_1, x_2, \ldots, x_n\}$ is the array of independent data; $\boldsymbol{y} = \{y_1, y_2, \ldots, y_n\}$ is the array of dependent data; $\bar{x}$ is the mean of $\boldsymbol{x}$; and $\bar{y}$ is the mean of $\boldsymbol{y}$.

An example of a linear transform function with is provided in Figure 11. This example transforms flow (cfs) to stage (ft) and has an intercept (α) of 10, a slope (β) of 0.5, and a standard error (σ) of 20.

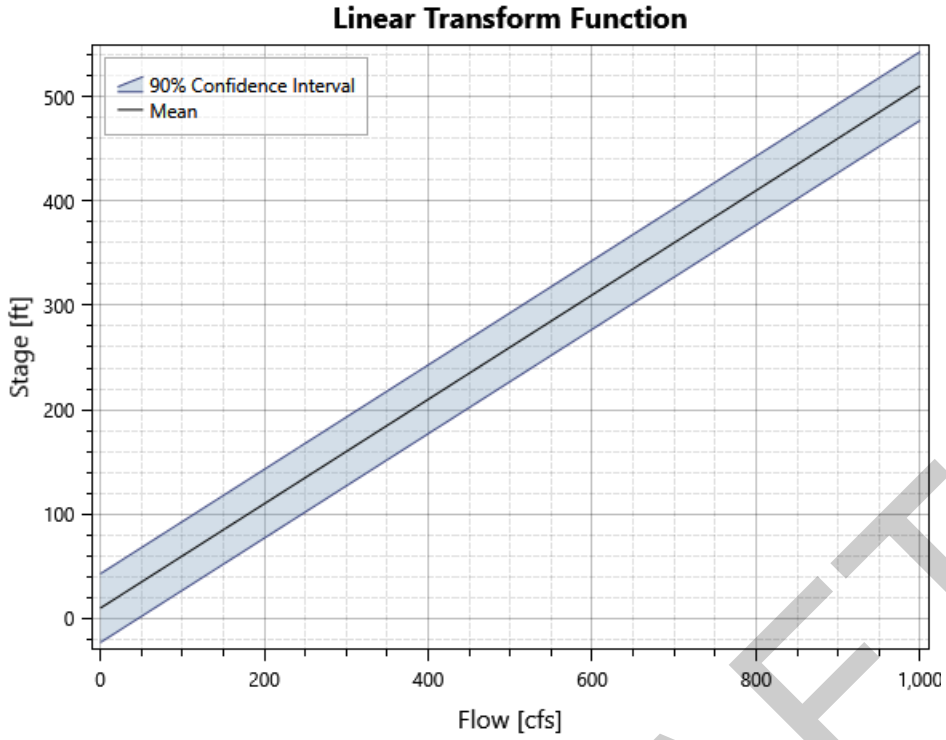


Figure 11 - Example of a linear transform function.

Power

Hazard levels can be transformed using a power function:

$$h(x) = y = \alpha(x - \xi)^{\beta} \cdot \varepsilon \qquad \textit{Equation 21}$$

where x is the hazard level; y is the transformed hazard level; α is the intercept coefficient; β is the exponent coefficient; ξ is the location; and ε is the model error, or residual, which is assumed to be log-normal and independently distributed (NID) with zero log-space mean and log-space standard error σ, or $\log \varepsilon \sim \Phi(0, \sigma)$.

To estimate the parameters for this function, the X and Y data should first be log-transformed so that the model becomes a simple linear equation:

$$\log y = \log \alpha + \beta \log(x - \xi) + \log \varepsilon \qquad \textit{Equation 22}$$

Then, the parameters can be estimated using OLS as described in the previous section. The value of both y and $(x - \xi)$ must be greater than zero. When simulating risk with full uncertainty, power transform functions are sampled from a Normal distribution using Equation 22 with Algorithm 5 from the previous section. For example, a random power transform function at the 0.9 percentile is created as follows:

$$y_i = e^{\{\log \alpha + \beta \log(x_i - \xi) + \Phi^{-1}(0.9|0,\sigma)\}} \quad (i = 1,2,\cdots,n) \qquad \textit{Equation 23}$$

There is also an option to use the inverse of Equation 21. For example, a flow-stage rating curve might be estimated with stage as the x variable and flow as the y. The inverse power transform function is as follows:

$$y = \left(\frac{x}{\alpha \cdot \varepsilon}\right)^{\frac{1}{\beta}} + \xi \qquad \textit{Equation 24}$$

As before, a random inverse power transform function is created as:

$$y_i = \left(\frac{x_i}{e^{\{\log \alpha + \Phi^{-1}(0.9|0,\sigma)\}}}\right)^{\frac{1}{\beta}} + \xi \quad (i = 1,2,\cdots,n) \qquad \textit{Equation 25}$$

An example of an inverse power transform function is provided in Figure 11. This example transforms flow (cfs) to stage (ft) and has an intercept coefficient (α) of 10, an exponent coefficient (β) of 2, a location coefficient (ξ) of 0, and a standard error (of log) (σ) of 0.1.

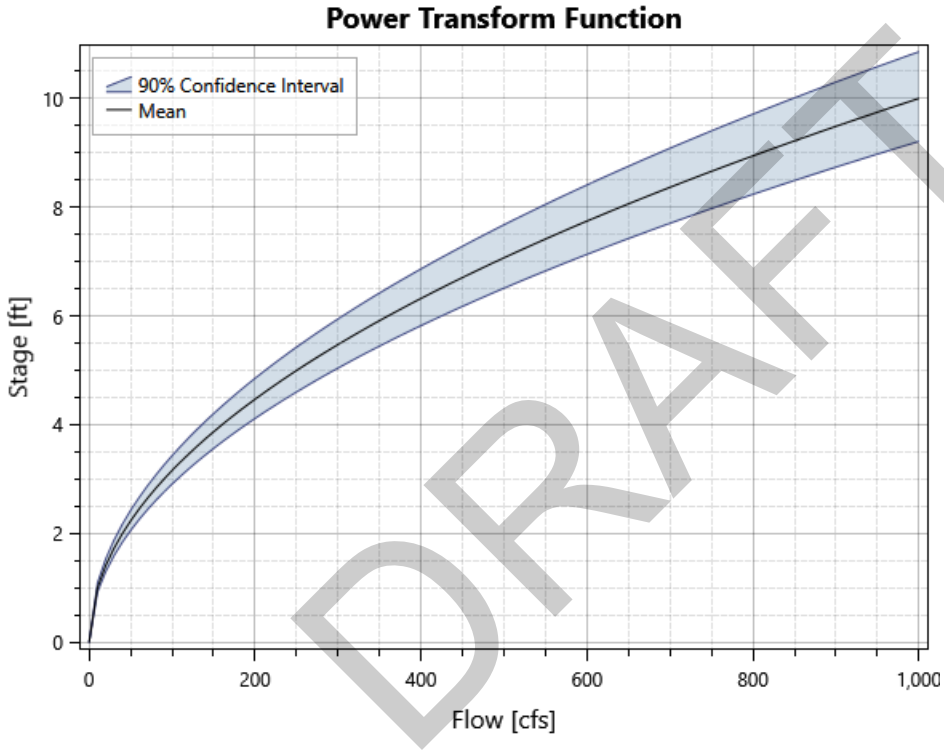


Figure 12 - Example of a power transform function.

Tabular

Hazard levels can be transformed using a tabular (or non-parametric) relationship of hazard levels and transformed hazard levels. The transformation is performed with linear interpolation:

$$h(x) = y = y_i + (y_{i+1} - y_i)\left(\frac{x - x_i}{x_{i+1} - x_i}\right) \qquad \textit{Equation 26}$$

where x is an array of hazard values $\boldsymbol{x} = \{x_1, x_2, \ldots, x_n\}$ for $x_i < x < x_{i+1}$; and y is an array of transformed hazard values $\boldsymbol{y} = \{y_1, y_2, \ldots, y_n\}$ for $y_i \leq y \leq y_{i+1}$. Extrapolation will not be performed for x values outside the user-defined range of hazard levels. A log-transform can be applied to both the x and y values to improve the accuracy of the interpolation.

A flow-stage rating curve will likely be derived by a hydraulic model, such as HEC-RAS. The modeled flow-vs-stage results can then be entered as tabular data into RMC-TotalRisk. An example of a tabular transform function is provided in Figure 13. Uncertainty can be defined for the transformed hazard for each tabular ordinate. A distribution must be selected to define uncertainty, and the parameters for the selected distribution must be entered for every ordinate in the tabular data. The uncertainty at each hazard level must be entered such that the confidence intervals are monotonically increasing with increasing hazard levels.

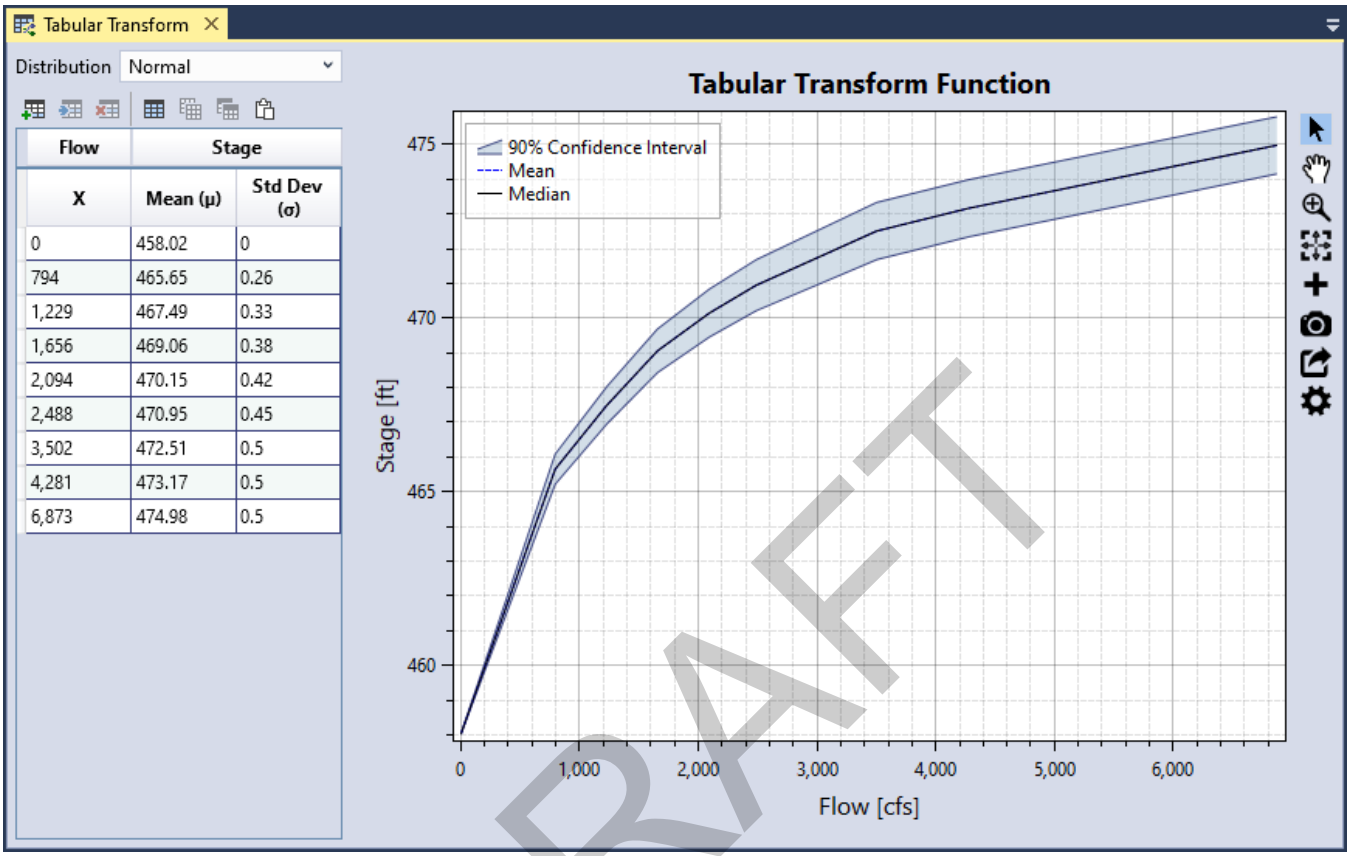


Figure 13 - Example of a tabular transform function.

When simulating risk with full uncertainty, tabular transform functions are sampled in the same manner as described in Algorithm 3. A new tabular transform function is created each Monte Carlo realization, which is then used to derive transformed hazard levels in the risk analysis.

System Response Functions

A system response function describes the conditional probability of failure of the system for various hazard levels, such as water surface elevations. System response functions are sometimes referred to as *fragility* curves.

Most engineering risk problems involve two opposing factors: a resistance (or capacity) R at hazard level x, and the load (or demand) S at hazard level x. When the load exceeds the resistance, failure occurs. It is important to note that he word “failure” does not necessarily imply fracture, breach, or collapse of the structure or system being evaluated. The term here is used in the general sense, meaning that the system fails to meet the demand placed on it. In engineering reliability analysis this is referred to as the *limit state*, which indicates general unsatisfactory performance, not just failure.

In engineering reliability analysis, the annual probability of failure p_F of a system is defined as the probability that the resistance is less than or equal to a specific load:

$$p_F = P(R \leq S) \qquad \textit{Equation 27}$$

$$p_F = \int_{-\infty}^{\infty} F_{R|S}(s|s) \cdot f_S(s) \cdot ds \qquad \textit{Equation 28}$$

where $F_{R|S}(s|s)$ is the conditional CDF of R given the demand $S = s$, and $f_S(s)$ is the marginal PDF of S, or hazard function distribution. For more details, the **R-S Reliability Formulation of Risk** section provides a formal derivation of Equation 28.

In RMC-TotalRisk, the system response function models the distribution of the resistance given a specific hazard level (or demand). Or stated another way, the response function is the CDF for the resistance of the system. The conditional probability of failure, or system response probability (SRP), is the probability that the resistance is less than or equal to a specific load:

$$P(F|x) = F_{R|S}(s|s) = P(R \leq S|S = s) \qquad \textit{Equation 29}$$

All system response functions in RMC-TotalRisk are modeled as a continuous random variable (see **Appendix A**). As such, each have a probability density function (PDF), $f(x)$, a cumulative distribution function (CDF), $F(x)$, a survival function, $S(x) = 1 - F(x)$, and an inverse CDF, $F^{-1}(p)$. However, there is an important distinction to be made.

In most risk analyses, the conditional probabilities of failure will be strictly increasing with increasing hazard levels. However, there are scenarios where the resistance of a dam or levee is really a multivariate function of different hazards. For example, consider a levee with flooding on the river side (headwater) and on the interior drainage side (tailwater). The height of both the headwater and tailwater will affect the probability of failure for an internal erosion failure mode. There is a possibility that the probability of failure can decrease with high river stages because tailwater on the interior of the levee is strongly positively correlated with the extreme riverine flooding. The higher tailwater levels increase the resistance and decrease the probability of failure for that type of failure mode.

Given the potential for these types of complex multivariate scenarios, and to allow for flexibility, RMC-TotalRisk does not require the probabilities of nonparametric system response functions to be strictly increasing. In addition, the probabilities do not need to be exhaustive, i.e., they do not need to be cumulative from 0 to 1.

System response functions can be defined as either a parametric or nonparametric distribution. A parametric distribution has a theoretical mathematical form, with input parameters, that describes how the probabilities are distributed over the values of a random variable. A nonparametric probability distribution makes very few assumptions about the underlying theoretical model, and it is based on empirical or simulated data. The following subsections describe the various response function options in RMC-TotalRisk.

Parametric

A response function can be defined with a parametric probability distribution. The following distribution options are available for parametric response functions:

- Exponential
- Gamma
- Logistic
- Ln-Normal
- Log-Normal (base 10)
- Normal
- Weibull

For details on these distributions, please see **Appendix E**. Parametric distributions can be estimated from distribution fitting software, such as *RMC-BestFit*, *HEC-SSP*, or other various commercial software.

As discussed in the parametric hazard function section, the parametric bootstrap [15] [16] is used to quantify uncertainty in the parametric response function, as described in Algorithm 1. The user must enter an effective record length (ERL), which is a measure of information content in the fit of the distribution. The longer the ERL, the less uncertainty and narrower the confidence intervals.

By default, the bootstrap analysis has 10,000 realizations because this is the maximum number of Monte Carlo realizations allowed in the risk analysis. When estimating risk with full uncertainty, one of the 10,000 bootstrapped parameter sets is selected at random each Monte Carlo realization (see Algorithm 2) Then, a new parametric probability distribution is created from the parameter set. The response probability given a hazard level can then be computed within the risk analysis for that realization.

An example of a parametric hazard is provided in Figure 5. This figure shows a Ln-Normal distribution for the probability of failure given annual maximum peak river stage (ft) with a mean of 10, standard deviation of 2, and an ERL of 30.

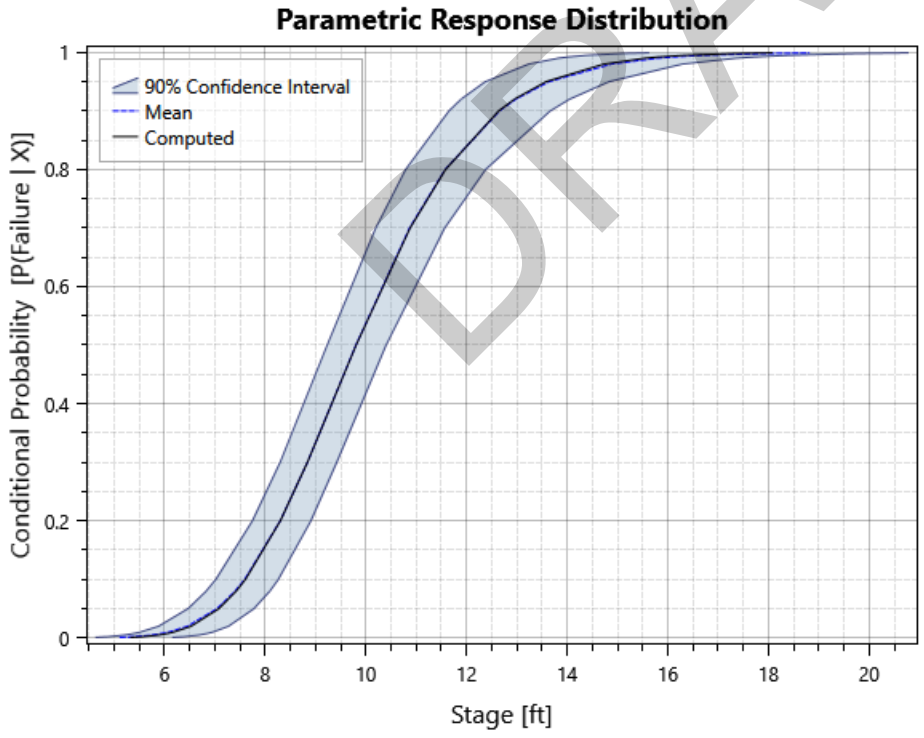


Figure 14 - Example of a parametric response function.

Nonparametric

Tabular

Response functions can be defined with a tabular (or nonparametric) relationship of hazard levels and conditional probabilities of failure. The tabular response function is computed using linear interpolation as follows:

$$P(F|x) = p_i + (p_{i+1} - p_i)\left(\frac{x - x_i}{x_{i+1} - x_i}\right) \qquad \textit{Equation 30}$$

where $P(F|x)$ returns the conditional probability of failure for a given hazard level x; and there is an array of continuous hazard values $\boldsymbol{x} = \{x_1, x_2, \dots, x_n\}$ for $x_i < x < x_{i+1}$with conditional probabilities $\boldsymbol{p} = \{p_1, p_2, \dots, p_n\}$ with $0 \leq p_i \leq 1$. The conditional probabilities do not need to be in any strict order, but it is typical that in most cases the probabilities will be increasing with increasing hazard levels. In addition, the probabilities do not need to be exhaustive, i.e., the do not need to be cumulative from 0 to 1.

There is often a need to apply a transform to the hazard and probability values to improve the accuracy of the linear interpolation. A log-transform can be applied to the $\boldsymbol{x}$ and/or $\boldsymbol{p}$ values. In addition, a Normal z transform can be applied to the $\boldsymbol{p}$ values.

Uncertainty can be defined for the probability of failure for each tabular ordinate. A distribution must be selected to define uncertainty, and the parameters for the selected distribution must be entered for every ordinate in the tabular data. An example of a tabular response function with uncertainty defined with a triangular distribution provided in Figure 15.

The uncertainty analysis for the tabular response function is performed as described in Algorithm 3. When simulating risk with full uncertainty, a new tabular response function is created each Monte Carlo realization, which is then used to derive the probability of failure given a hazard level in the risk analysis.

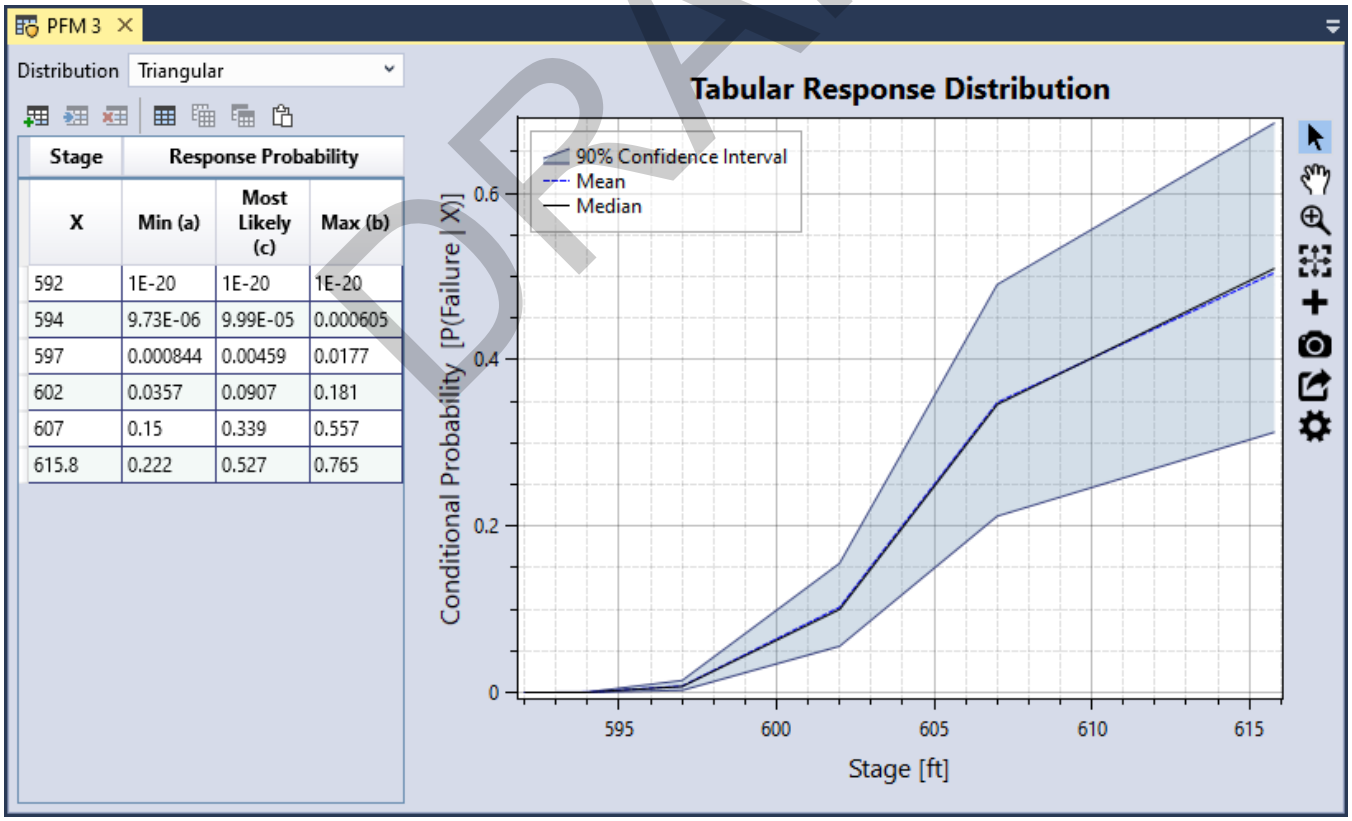


Figure 15 - Example of a tabular response function with uncertainty.

Bivariate

A bivariate response function provides a simple way to define a tabular response function that is conditional on two hazards. The response probability of a bivariate response function is derived from the total probability theorem (see **Appendix A**).

$$P(F|x) = \sum_{i=1}^{n} P(F|x, y_i) \cdot P(y_i) \qquad \textit{Equation 31}$$

where $P(F|x)$ is the conditional probability of failure given the primary hazard level x; $P(F|x, y_i)$ is the conditional probability of failure given the primary hazard level x and secondary hazard y_i; and $P(y_i)$ is the likelihood of the secondary hazard level y_i. The summation in Equation 31 can be interpreted as a weighted sum, and as such $P(F|x)$ is sometimes called the *expected* or *total* probability.

The bivariate response function is best understood through an example. Seismic failure modes for dams are often conditional on the water surface elevation (or stage) in the reservoir when the earthquake occurs, and the peak ground acceleration (PGA) of the earthquake. Table 2 shows the nonexceedance probabilities for various PGA hazard levels. Table 3 shows an example bivariate response table for the seismic failure mode. In the case, the primary hazard is reservoir stage (ft). The secondary hazard is the PGA, which is usually expressed in terms of gravity, or *g* [5]. The likelihood (or weight) of each PGA scenario is provided Table 4. As shown in Table 4, the weights are computed as the probability of being within an interval $P(a \leq X \leq b)$.

Table 2 – Nonexceedance probabilities for peak ground accelerations.

Peak Ground Acceleration (g)	$P(X \leq x)$
0.2	0.9989760
0.3	0.9995930
0.4	0.9997920
0.5	0.9998830
0.6	0.9999308
0.7	0.9999579
0.8	0.9999736
0.9	0.9999832
1.0	0.9999893

Table 3 – Example bivariate response table for a seismic failure mode.

		Peak Ground Acceleration (g)			
		0.2	0.4	0.6	0.8
Stage (ft)	2560.0	6.00E-09	3.60E-06	3.06E-05	5.39E-05
	2585.5	3.10E-06	1.86E-03	1.58E-02	2.78E-02
	2590.0	1.13E-05	6.75E-03	5.74E-02	1.01E-01
	2605.5	4.58E-05	2.75E-02	2.34E-01	4.11E-01
	2611.0	7.20E-05	4.32E-02	3.67E-01	6.47E-01
	2633.5	9.07E-05	5.44E-02	4.62E-01	8.14E-01

Table 4 - Likelihoods (or weights) of secondary hazard levels.

Secondary Hazard Levels			
PGA (g)	Range	$P(a \leq X \leq b)$	Weight
0.2	0.0 to 0.3	$P(0 \leq X \leq 0.3) = 0.999593$	9.99593E-01
0.4	0.3 to 0.5	$P(0.3 \leq X \leq 0.5) = 0.999883 - 0.999593 = 0.00029$	2.90000E-04
0.6	0.5 to 0.7	$P(0.5 \leq X \leq 0.7) = 0.9999579 - 0.9998830 = 0.0000749$	7.49000E-05
0.8	>0.7	$P(0.7 \leq X \leq \infty) = 1 - 0.9999579 = 0.0000427$	4.21000E-05

Each column in Table 3 can be viewed as a single tabular response function, $F_i(\cdot)$, one for each PGA level i. Then, Equation 31 can be rewritten as a mixture distribution as follows:

$$P(F|x) = \sum_{i=1}^{n} \omega_i \cdot F_i\,(x) \qquad \textit{Equation 32}$$

where $F_i(\cdot)$ is the response function for each secondary hazard level i; and ω_i is the weight or likelihood of the secondary hazard level i, with $0 \leq \omega_i \leq 1$ and $\sum_{i=1}^{n} \omega_i = 1$.

The bivariate response function from Table 3 is shown in Figure 16. Weights can be entered manually or automatically derived based on the selected secondary hazard function. Uncertainty cannot be modeled with the bivariate response function. If there is a desire to include uncertainty, the composite response function should be used instead.

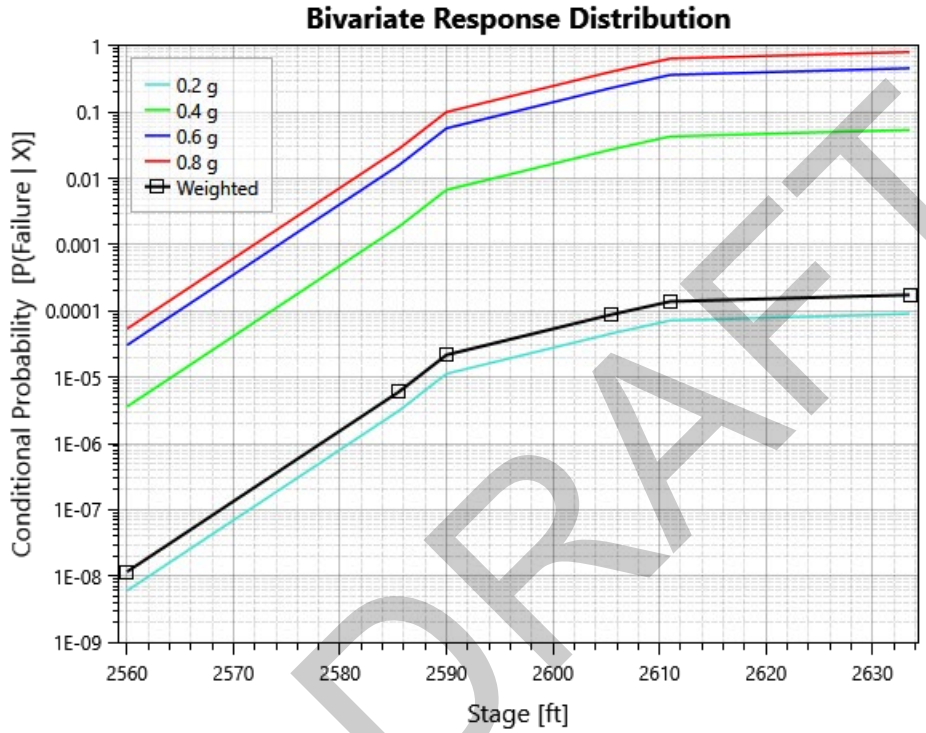


Figure 16 - Example of bivariate response function for a seismic failure mode.

Event Tree

A response function can be defined from an event tree. Event tree analysis (ETA) represent the logic of how an initiating event, like a flood or earthquake, can lead to various types of damage and failure [30]. It is common practice to develop detailed event trees for individual PFMs to clearly identify the full sequence of steps required to obtain failure or breach. Each identified PFM is decomposed into a sequence of component events and conditions that must occur for there to be a failure. This ensures that due consideration is given to each event in the failure sequence [5].

Terminology

This section provides definitions for the event tree terminology used commonly in quantitative risk analyses. Similar terms and definitions are provided in [5] and [30].

- **Node**: A branching point in the event tree that signifies a random event in the event tree. This is also commonly referred to as an *event* or *state*. There are four basic node types that define the event tree:

  Initiating Hazard Node: Always the first node in an event tree. This node defines the hazard levels. It is crucial to relating system response probabilities to the hazard function. For a more accurate assessment of risk, probabilities should be defined for enough hazard levels to cover nearly the entirety of probability space. Only hazard levels are entered for this node, not hazard exceedance probabilities.

  Chance Node: Represents the probability that the given event will occur for each hazard level defined in the initiating hazard node. Probabilities can be defined as a single value for all hazard levels (Single Value), unique for each hazard level (Multi Value), or from another source (see reference node below). This is the most fundamental component of an event tree.

  Reference Node: The reference node performs the same task as a chance probability node except that instead of probability being defined at the node itself, it is defined using a previously added response function or node in the event tree. The selected function or node can be sampled independently or the same for all references to the source.

  Remainder Node: This node represents the probability that remains when all other chance node and reference node probabilities for the current branching point are considered. For example, from a given event with two potential outcomes the remainder node represents the probability that neither of the potential outcomes occurs. The remainder probability is computed automatically and is not set by the user.

- **Branch**: The line that connects two nodes together.

- **End Node**: A node that has no downstream branches. The end node defines the end state for a sequence of events. This is also commonly referred to as a *leaf node* or *terminal node*.

- **Pathway**: A unique sequence of events representing a possible failure progression. The probability of this pathway occurring is computed as the joint probability of each node in the series that connects the initiating hazard node to the end node. This is also commonly referred to as *path*, *sequence*, *connections*, or *root-to-node*.

- **Upstream Nodes**: All nodes to the left of a selected node in the tree. The upstream nodes of the selected node must occur prior to the selected node event. This is also commonly referred to as *parent node*, *conditional events*, or *preceding nodes*.

- **Downstream Nodes**: All nodes to the right of a selected node in the tree. The downstream nodes occur after the selected node event. This is also commonly referred to as *child node*, *conditional events*, or *proceeding nodes*, or *subsequent nodes.*

- **Node Probability**: The probability that the selected node event occurs conditioned on the occurrence of the upstream nodes in the tree.

- **Event Likelihood**: The likelihood, or probability, of a node occurring in the event tree for a given hazard level. This is the joint probability of the selected node event and all the upstream nodes occurring.

Event Probability

A conceptual event tree is shown below in Figure 17. A sequence of events is characterized in an ETA as nodes connected by branch lines where each node represents an event or state. The sequence of events along a branch, going from left to right, is referred to as a pathway. All nodes can have multiple downstream nodes, each with their own branches and pathways. In Figure 17, the sequence $\{x, P_1, P_2, P_3, P_4, P_5\}$ represents a pathway to failure.

If a node does not have any downstream nodes, it is considered a leaf node and represents the end state of the tree, e.g., breach or component failure leading to consequences. In Figure 17, the chance node p_5 is a leaf node representing the pathway to failure. Each remainder node in this example represents a pathway to non-failure. The remainder node r_5 represents a leaf node for non-failure.

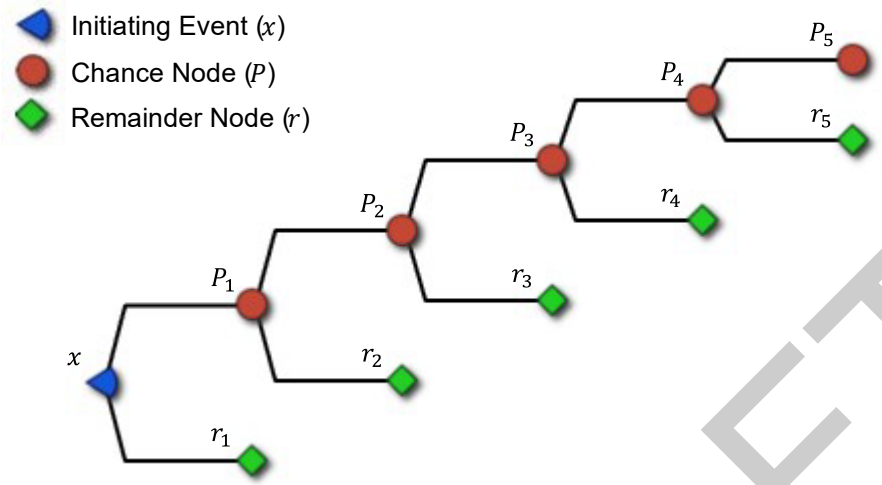


Figure 17 - Conceptual event tree diagram.

All nodes, except the initiating event node, have a single upstream node. An upstream node defines the event that must occur prior to occurrence of the selected branch node. Probability estimates can be uniquely specified at each node (typically from expert elicitation), reference another node in the event tree, or reference another system response function in the TotalRisk project. Nodal probabilities are conditional estimates given the preceding event has occurred and are defined as follows:

$$P_1 = P(E_1|x) \qquad \textit{Equation 33}$$

where P_1 is the probability of event E_1 occurring given the initiating event x. Similarly, the probability of event E_2 is:

$$P_2 = P(E_2|x, E_1) \qquad \textit{Equation 34}$$

so on and so forth. The probability of failure $P(F|x)$ is then estimated as the joint probability of all events occurring in the sequence $\{x, P_1, P_2, P_3, P_4, P_5\}$:

$$\begin{aligned} P(F|x) &= P(E_1 \cap E_2 \cap E_3 \cap E_4 \cap E_5) \\ &= P(E_1|x) \cdot P(E_2|x, E_1) \cdot P(E_3|x, E_1, E_2) \cdot P(E_4|x, E_1, E_2, E_3) \\ &\cdot P(E_5|x, E_1, E_2, E_3, E_4) \end{aligned} \qquad \textit{Equation 35}$$

which is known as the *multiplication rule* or *chain rule* for probability (see **Appendix A**). Equation 35 can be written more generally as:

$$P(F|x) = P\left(\bigcap_{i=1}^{n} E_i|x\right) = \prod_{i=1}^{n} P_i|x \qquad \textit{Equation 36}$$

Each pathway j downstream of a node must be mutually exclusive and collectively exhaustive. Figure 18 shows an example where there are three pathways to failure after event E_1, which are highlighted with the red box.

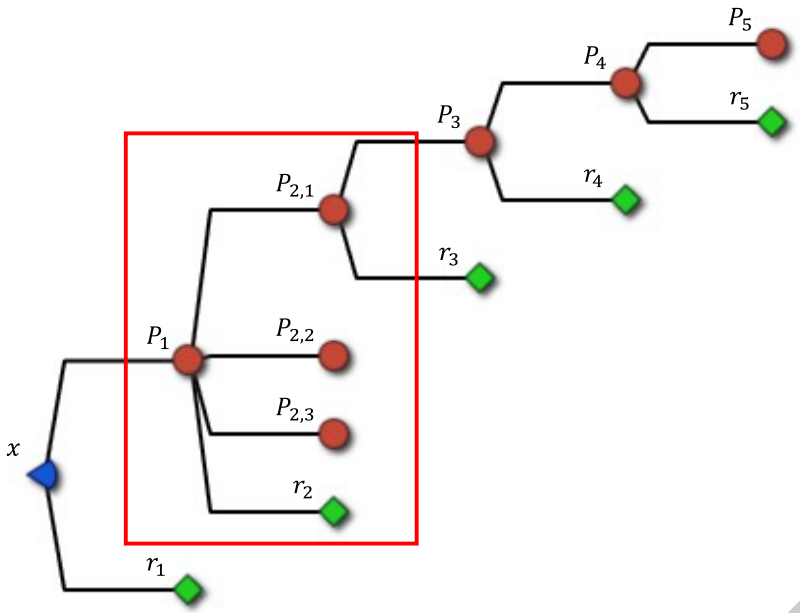


Figure 18 - Event tree diagram showing an example of multiple downstream nodes and possible pathways.

Following from the *addition rule* (see **Appendix A**), the probability that any of the events occur downstream of a node is the sum of each branch probability originating from the node. The sum of branch node probabilities must not exceed one. The remaining probability r from any node i is the probability that none of the proceeding events occur, also called the compliment or probability of non-occurrence. The following equation is used to automatically calculate the remaining probability:

$$r_i|x = 1 - P\left(\bigcup_{j=1}^{m} E_{i,j}|x\right) = 1 - \sum_{j=1}^{m} P_{i,j}|x \qquad \textit{Equation 37}$$

where r_i is the probability that no proceeding event nodes occur after event node i; and $\boldsymbol{P}_i = \{P_{i,1}, P_{i,2}, \cdots P_{i,m}\}$ is a vector of node probabilities from each branch originating from node i for hazard level x.

Response Function

As discussed above, the likelihood of a node event occurring is conditional on the occurrence of all preceding events in the tree. Because the node probabilities are conditional on the preceding events occurrence, the probability of a node occurring given the initiating event for any node is the product of all node probabilities connecting the initiating event and target node. In other words, since each node event is conditional on the preceding event occurring, the probability of any node event occurring is the product of each probability in sequence from the initiating event to the event in question. Therefore, the probability of failure for a pathway j for an initiating event x_k is computed as:

$$P(F|x_k)_j = P\left(\bigcap_{i=1}^{n} E_i|\, x_k\right) = \prod_{i=1}^{n} P_{i,j}|\, x_k \qquad \textit{Equation 38}$$

where x_k is the initiating event (or hazard level); and $\boldsymbol{P}$ is a vector of node probabilities connecting the initial event x_k to node event i moving from left to right on the pathway j. An example calculation is shown in Figure 19 below.

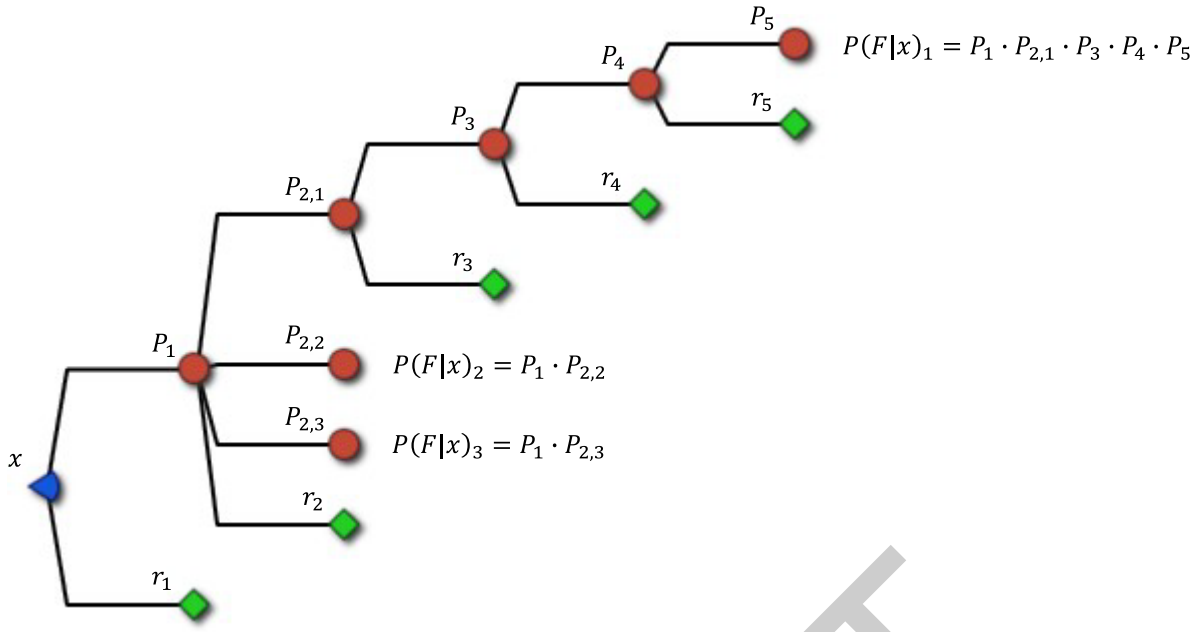


Figure 19 - Event tree diagram showing the calculation for likelihood of a node occurring given the initiating event.

The total probability of failure given an initiating event x_k is the sum of the likelihood of each chance leaf node occurring. Each chance leaf node represents the end state (e.g., breach, component failure, etc.) that leads to consequences. The probability of failure, or system response probability (SRP), for initiating event x_k is given by:

$$P(F|x_k) = P\left(\bigcup_{j=1}^{m} E_{i,j}|x\right) = \sum_{j=1}^{m} P(F|x_k)_j \qquad \textit{Equation 39}$$

A generalized equation can be defined by combining Equation 38 and Equation 39 using the *sum-product* formulation:

$$P(F|x_k) = P\left(\bigcup_{j}\bigcap_{i} E_{i,j}|x_k\right) = \sum_{j}\prod_{i} P_{i,j}|x_k \qquad \textit{Equation 40}$$

The system response is evaluated over a range of monotonically increasing hazard levels (initiating events), such as water surface elevations. The range of hazard levels and associated SRPs are treated as a tabular (nonparametric) response function. An example of a system response function derived from an event tree is shown in Figure 20 below.

Uncertainty Analysis

The probability values assigned at chance nodes can be defined with uncertainty. Monte Carlo sampling is used to account for the uncertainty in node probabilities. Random samples are defined for each chance node in the event tree. If the event tree contains reference nodes and the reference node is set to be sampled independently, that reference node will get a unique sample. Otherwise, the reference node will receive the same random sample as the referenced function. The probability distribution, or reference function, at each node is then sampled to get discrete probabilities at each node for each hazard level defined. Once each node has a discrete probability sampled, the system response probability is calculated using Equation 40 above at each hazard level to define the system response function.

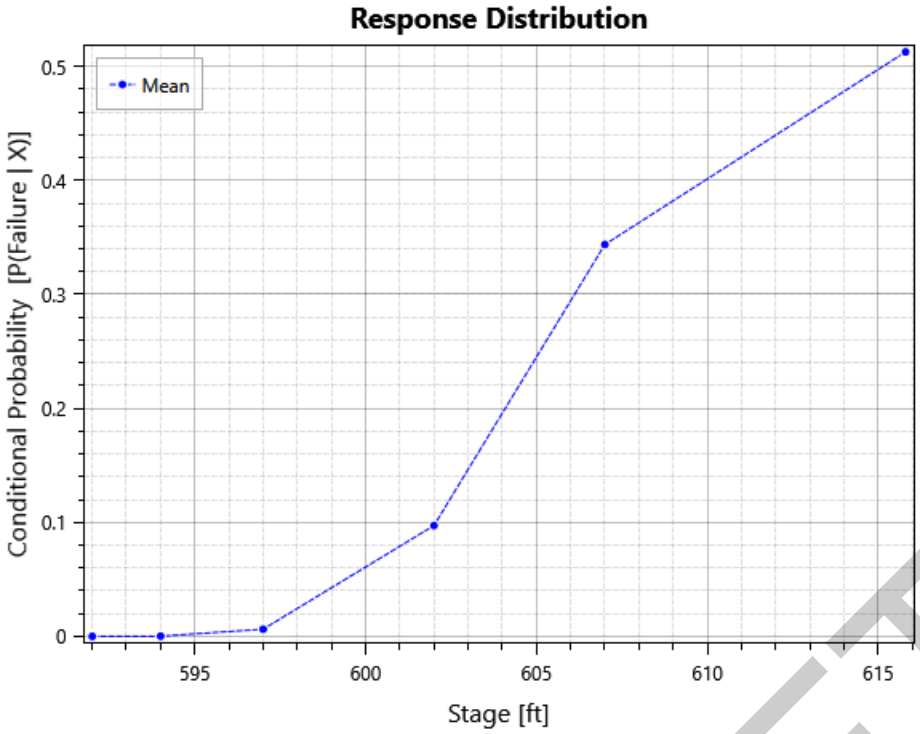


Figure 20 - Example system response function in RMC-TotalRisk.

Since each chance node in the event tree can have a probability distribution, there is the possibility that the branch node probabilities will sum greater than one during the Monte Carlo simulation. If this occurs, the branch node probabilities are normalized to sum to exactly one. If during the uncertainty analysis normalization is required, a warning message will be sent to the user. If normalization is required during the simulation, the following equation is used to automatically calculate the remaining probability:

$$r_i|x = 1 - \sum_j z_{i,j}\,|x \qquad \textit{Equation 41}$$

where r_i is the probability that no proceeding event nodes occur after event node i; and $\boldsymbol{z}_i = \{z_{i,1}, z_{i,2}, \cdots z_{i,m}\}$ is a vector of normalized node probabilities from each branch originating from node i. The following equation is used to normalize multiple branch probabilities:

$$z_{i,j} = P_{i,j}, \qquad \sum_j P_{i,j} \leq 1$$

$$z_{i,j} = \frac{P_{i,j}}{\sum_j P_{i,j}}, \qquad \sum_j P_{i,j} > 1 \qquad \textit{Equation 42}$$

where $z_{i,j}$ is the normalized node probability and $\boldsymbol{P}_i = \{P_{i,1}, P_{i,2}, \cdots P_{i,m}\}$ is a vector of unadjusted probability values sampled at each branch originating from a single event node i with $0 \leq P_{i,j} \leq 1$. An example of the remaining probability calculation for both conditions in Equation 42 above is shown in Figure 21 below.

$\mathbf{0.3 + 0.4 + 0.8 = 0.8 \leq 1}$ $\mathbf{0.6 + 0.4 + 0.3 = 1.3 > 1}$

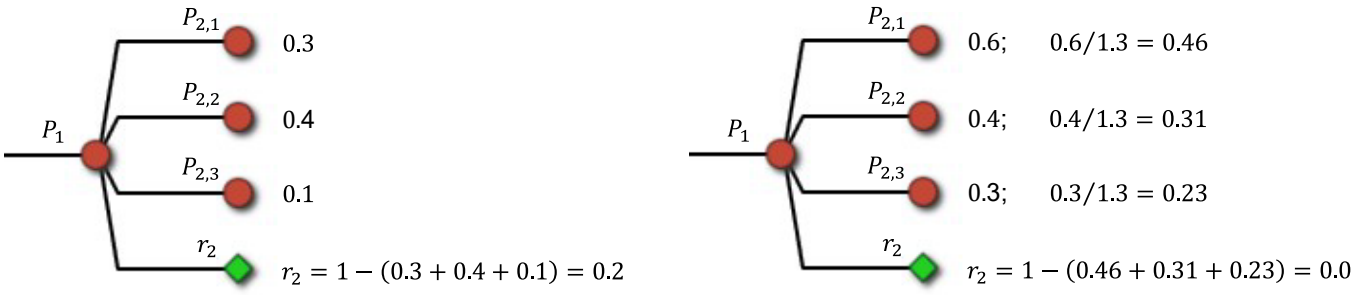


Figure 21 - Event tree diagram showing remainder probability calculations with multiple branch node probabilities that sum to less than and greater than one.

An algorithm for calculating the system response function with confidence intervals is shown below. This algorithm describes the routine given a single hazard level of interest. It can easily be extended to include a list of hazard levels to evaluate. Figure 22 shows an example of a system response function with confidence intervals derived from an event tree uncertainty analysis.

Algorithm 6 – Event Tree Response Uncertainty Analysis

$R \leftarrow$ number of Monte Carlo realizations
$n \leftarrow$ number of chance nodes
$x \leftarrow$ hazard level to evaluate
for $i \leftarrow 1$ to R **do**
  Sample uniformly at random with replacement an integer seed $s_i \sim \mathrm{U}(0,2147483647)$
  Create random number generator with seed $rnd(s_i)$
  $P(F|x)_i^* \leftarrow 1$
  for $j \leftarrow 1$ to n **do**
    Using rnd, sample uniformly at random with replacement a percentile $r_j \sim \mathrm{U}(0,1)$.
    if node j is a reference node and dependent **then** percentile equals source percentile
    Generate node probability $P_j^* \leftarrow F^{-1}(r_j|\theta)$ for node j
    $P(F|x)_i^* \leftarrow P(F|x)_i^* \cdot P_j^*$
  end for
end for
Estimate confidence intervals and mean probability for x from $\{P(F|x)_1^*, \cdots, P(F|x)_R^*\}$.

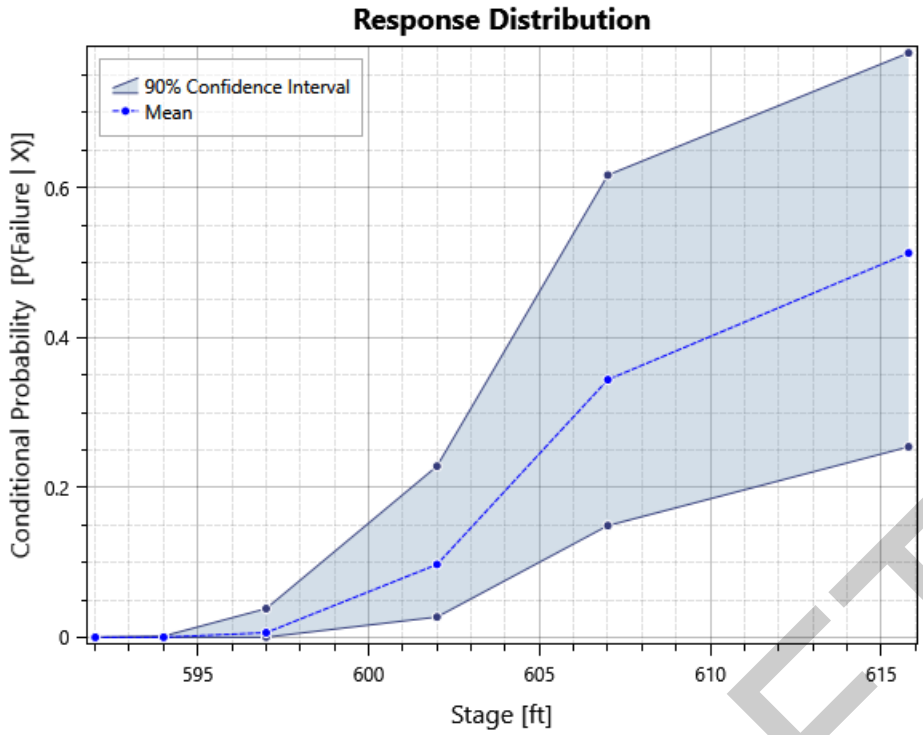


Figure 22 - Example system response function with confidence intervals in RMC-TotalRisk.

Diagnostics

There are several diagnostic tools in RMC-TotalRisk for performing sensitivity analysis on an event tree. The diagnostic tools provide a better understanding of how each node in the event tree is impacting the overall SRP for a given hazard level. Most of the diagnostic tools are only applicable when uncertainty exists in the nodal probability estimates. For the sensitivity analysis, 1,000 Monte Carlo realizations are computed by default. For more detail on sensitivity analysis in RMC-TotalRisk, see **Appendix G**. Diagnostic tools include comparing node likelihoods, correlating sampled node probabilities to SRP, and sensitivity of sampled node probabilities to SRP.

Event Likelihood

RMC-TotalRisk reports the likelihood of each node in the event tree occurring given an initiating hazard level. Knowing the likelihood of each node event provides a fuller picture of the sequence of events that must happen for the end state to occur. The likelihood of any node event occurring is discussed above in Equation 38. A five number summary of event likelihood is calculated for every node in the event tree for all Monte Carlo realizations. The five number summary consists of the sampled Minimum, 25th %-ile, Median, 75th %-ile, and Maximum event likelihoods for a given hazard level. As shown in Figure 23, each node is ranked in decreasing order of median event likelihood.

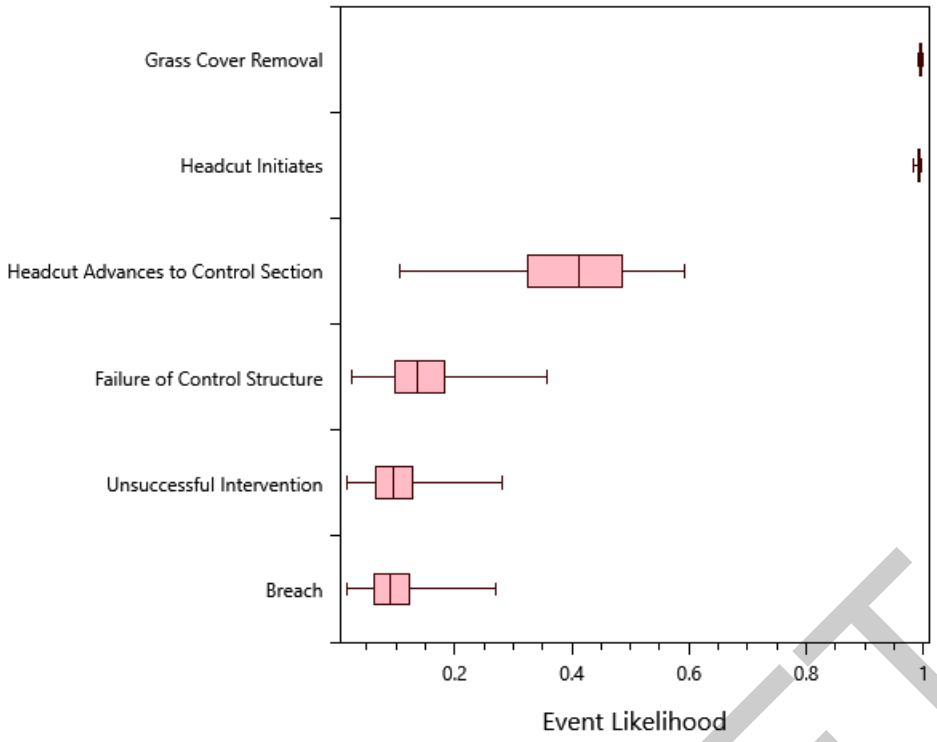


Figure 23 - Example box and whisker plot showing the likelihood of each chance node occurring in the event tree for a given hazard level.

Correlation

Correlating sampled node probabilities to the overall SRP in a Monte Carlo simulation provides a measure of which nodes are most associated with the output SRP for a given hazard level. This diagnostic tool is only available if uncertainty exists in the event tree chance nodes.

Pearson's correlation coefficient is used for computing the correlation between the nodes and the overall SRP. The correlation coefficient will be high if the sampled node probabilities are highly associated with the SRPs, i.e., if an input and output tend to move together.

The Pearson correlation coefficient is a value between -1 and 1. The sign of the coefficient indicates whether the association between the input and output is positive or negative. For a positive correlation, if the input value increases from the mean, the output value tends to also increase. For a negative correlation, if the input values increased from the mean, the output value tends to decrease. For example, chance nodes will have a positive correlation with the SRP, whereas remainder nodes will have a negative correlation. The size of the correlation coefficient indicates the strength of the association. Please see **Appendix G** for more details.

As shown in Figure 24, each chance node is ranked from largest to smallest correlation coefficient. This type of plot is commonly referred to as a *tornado plot*.

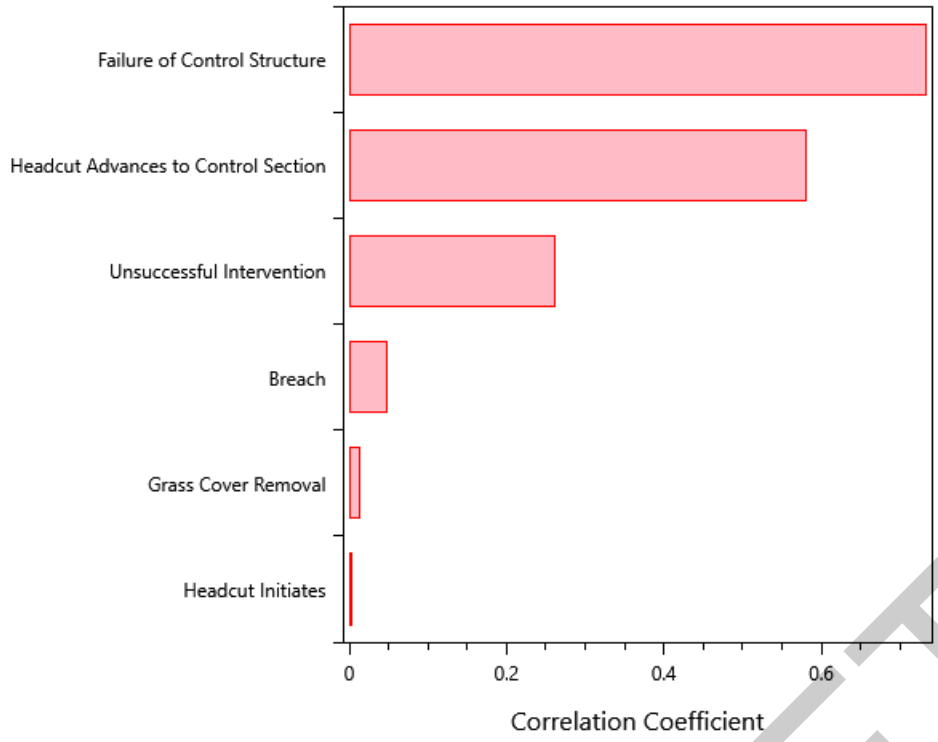


Figure 24 – Example bar chart showing correlation coefficients of sampled node probabilities in the event tree to the overall SRP for a given hazard level.

Sensitivity Index

Another useful diagnostic is the sensitivity index. The sensitivity index is a measure of the contribution to the variance that a single input node has on the output SRP. The sensitivity index is a value between 0 and 1, where 0 indicates a 0% contribution and 1 indicates a 100% contribution to variance. This diagnostic tool is only available if uncertainty exists in the event tree chance nodes.

A high sensitivity index indicates that reducing the uncertainty in the node probability will have a large reduction in the uncertainty of the overall SRP. The sensitivity index provides a measure for identifying which nodes to focus on to reduce uncertainty in the SRP function. Please refer to **Appendix G** for more information on how the sensitivity index is computed.

As shown in Figure 25, each node is ranked from largest to smallest sensitivity index. It can be seen from Figure 24 and Figure 25, that both the correlation coefficient and sensitivity index will typically provide similar rankings of chance nodes.

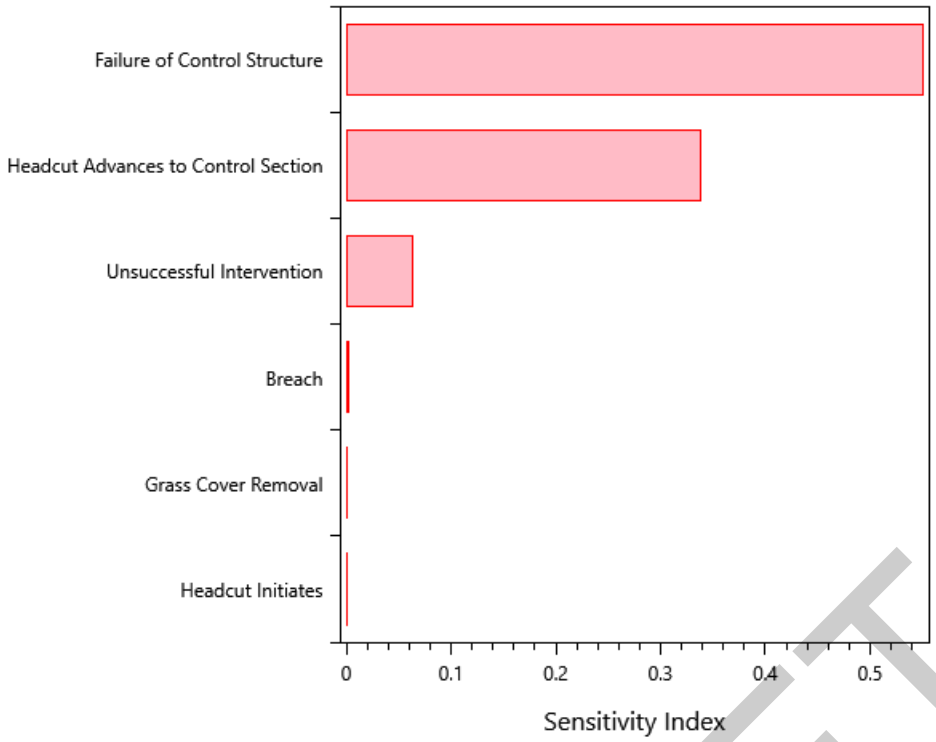


Figure 25 - Example bar chart showing sensitivity indices of nodes in the event tree to the SRP for a given hazard level.

Composite

A composite response function can be created by assigning weights (or likelihoods) to a list of response functions as follows:

$$F(x) = \sum_{i=1}^{n} \omega_i \cdot F_i(x) \qquad \textit{Equation 43}$$

where $F_i(\cdot)$ is the CDF for response function i; and ω_i is the weight or likelihood of hazard function i, with $0 \leq \omega_i \leq 1$ and $\sum_{i=1}^{n} \omega_i = 1$. This type of composite function is traditionally referred to as a *mixture distribution* [27].

Rather than using weights, a composite response function can alternatively be created by combining a list of response functions using the probability of union, assuming statistical independence between functions as follows:

$$F(x) = 1 - \prod_{i=1}^{n} [1 - F_i(x)] \qquad \textit{Equation 44}$$

where $F(\cdot)$ is the CDF of the composite response function; and $F_i(\cdot)$ is the CDF for response function i. Please see **Appendix A** for more details on the probability of union. Figure 26 shows an example of a composite response function for combining multiple failure modes for a levee.

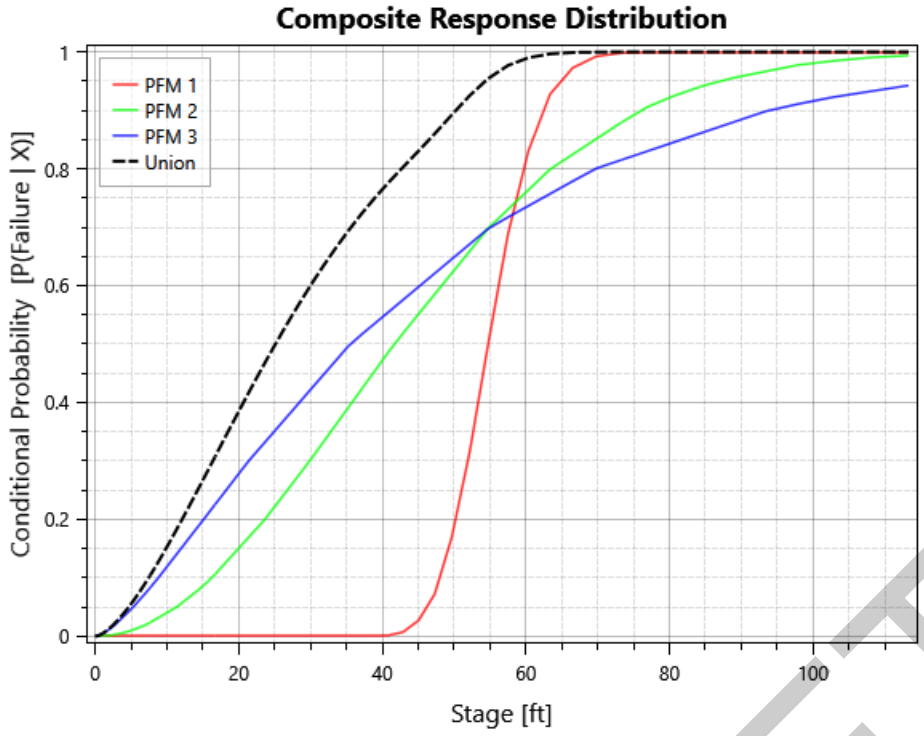


Figure 26 - Example of a composite response function for combining multiple failure modes using the probability of union.

The uncertainty analysis for the composite hazard functions is performed as described in Algorithm 6. This algorithm describes the routine given a single hazard level of interest. It can easily be extended to include a list of hazard levels to evaluate. When simulating risk with full uncertainty, a new composite response function is created each Monte Carlo realization, which is then used to derive probabilities of failure in the risk analysis.

Algorithm 6 – Composite Response Function Uncertainty Analysis

$R \leftarrow$ number of Monte Carlo realizations
$n \leftarrow$ number of response functions
$x \leftarrow$ hazard level to evaluate
for $i \leftarrow 1$ to R **do**
  Sample uniformly at random with replacement a percentile $r_i \sim \mathrm{U}(0,1)$.
  if mixture **then** $p_i^* = 0$ **else** $p_i^* = 1$
  for $j \leftarrow 1$ to n **do**
    Sample a random response function F_j^* given r_i
    if mixture **then** $p_i^* \leftarrow p_i^* + \omega_j \cdot F_j^*(x|r_i)$ **else** $p_i^* \leftarrow p_i^* \cdot \left[1 - F_j^*(x|r_i)\right]$
  end for
  if not mixture **then** $p_i^* \leftarrow 1 - p_i^*$
end for
Estimate confidence intervals and mean probability for x from $\{p_1^*, \cdots, p_R^*\}$.

Consequence Functions

A consequence function describes the consequences of failure or non-failure for various hazard levels, such as water surface elevations. Consequence functions are also sometimes referred to as *damage* functions [7]. All consequence functions in RMC-TotalRisk are nonparametric. The following subsections describe the various consequence function options in RMC-TotalRisk.

Tabular

Consequence functions can be defined with a tabular (or nonparametric) relationship of hazard levels and consequences. The consequences are computed with linear interpolation:

$$C(x) = y_i + (y_{i+1} - y_i)\left(\frac{x - x_i}{x_{i+1} - x_i}\right) \qquad \textit{Equation 45}$$

where x is an array of hazard values $\{x\} = \{x_1, x_2, \ldots, x_n\}$ for $x_i < x < x_{i+1}$; and y is an array of consequence values $\{y\} = \{y_1, y_2, \ldots, y_n\}$. The consequences values do not need to be in any strict order. A log-transform can be applied to both the x and y values to improve the accuracy of the interpolation.

The consequence values can be deterministic or defined with uncertainty by selecting a distribution. An example tabular consequence function with uncertainty is shown in Figure 27. The parameters for the selected distribution must be entered for every ordinate in the tabular data.

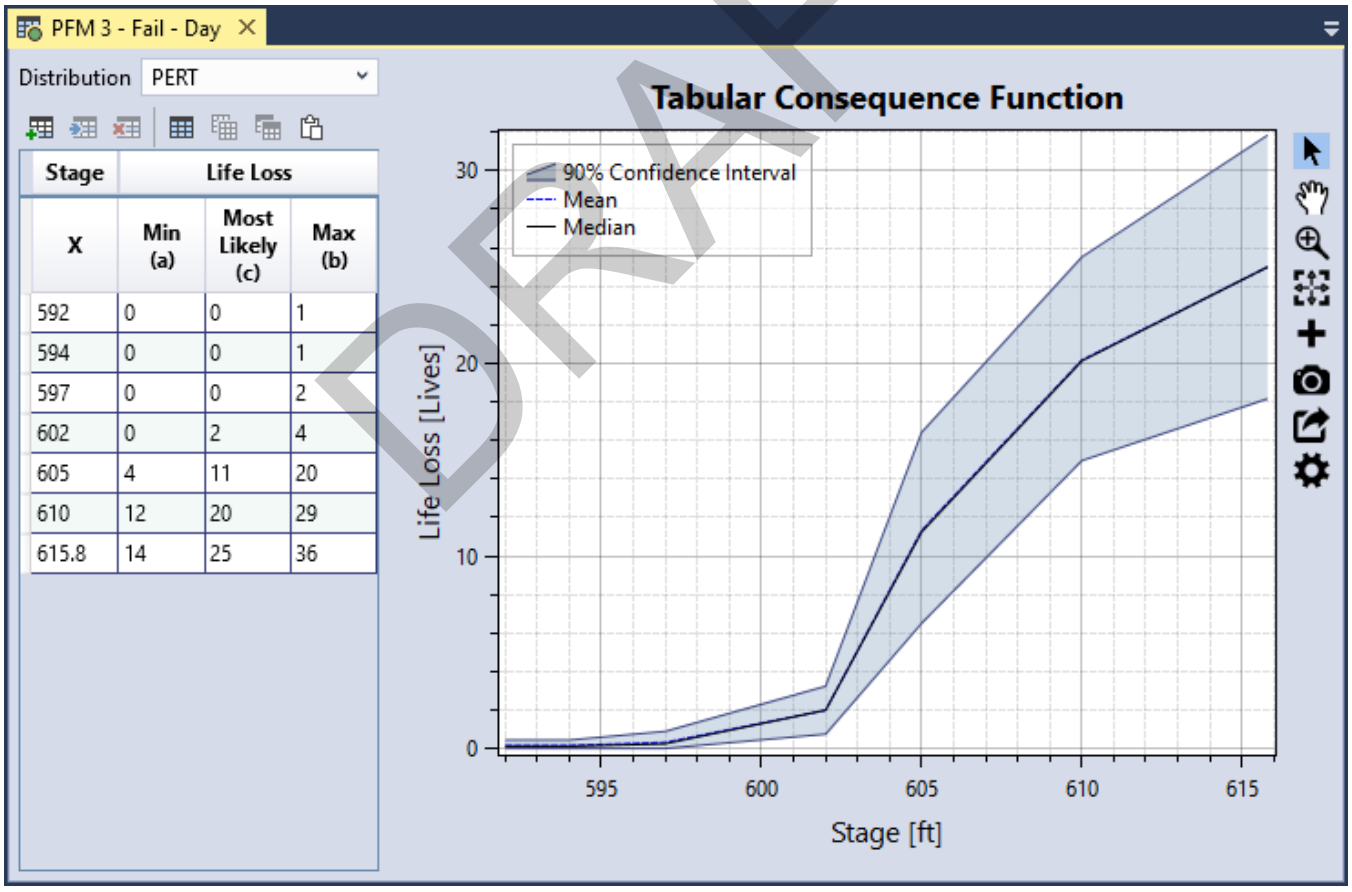


Figure 27 - Example of a tabular consequence function with uncertainty.

The uncertainty analysis for the tabular consequence functions is performed in the same manner as the other tabular functions in RMC-TotalRisk. When simulating risk with full uncertainty, a new tabular consequence function is created each Monte Carlo realization, which is then used to derive fail or non-fail consequences in the risk analysis.

Incremental consequences are the incremental losses or damages that failure might inflict over and above any damage which would otherwise occur for the same hazard event assuming the structure does not fail. Incremental consequences for a failure mode are the difference between the consequences of failure and non-failure for the same hazard level:

$$C_{\Delta}(x) = C_F(x) - C_{NF}(x) \qquad \textit{Equation 46}$$

where $C_{\Delta}(\cdot)$ are the incremental consequences; $C_F(\cdot)$ are the consequences of failure; and $C_{NF}(\cdot)$ are the consequences of non-failure.

In the Monte Carlo simulation, the consequence functions for fail and non-fail are sampled together as perfectly correlated within each failure mode. This is done to satisfy the definition that the consequences of failure and non-failure are associated with the same hazard event. In addition, if for any reason, negative incremental consequences are calculated in the risk simulation, they are set to zero and a warning message is provided to the user.

LifeSim

A tabular consequence function can be generated using the results from the consequence estimation software, *LifeSim*. In this case, each ordinate of the tabular function has consequence estimates based on results from a LifeSim simulation.

The RMC in association with HEC developed a modern implementation of *LifeSim* for the primary purpose of estimating life loss and economic damages from flood events. The key component of the *LifeSim* methodology is that the magnitude of life loss depends on whether people evacuate successfully and whether those who fail to evacuate can find adequate shelter [31]. *LifeSim* explicitly models the warning and evacuation of people during a flood and predicts the spatial distribution of fatalities within buildings and on roads. It uses an agent-based approach to track individuals throughout the warning and evacuation process. Although *LifeSim* was developed for dam and levee safety analyses, the software is not limited to breach flood hazards [32].

LifeSim employs Monte Carlo sampling techniques to capture uncertainty in various parameters that influence loss of life from flooding resulting in a distribution of potential outcomes as shown in Figure 28. Many model parameters can be defined with uncertainty, including but not limited to structure stability criteria, hazard identification time, hazard communication time, warning diffusion speed, protective action initiation timelines, and fatality rates for those that are inundated. Details on the methodologies applied in *LifeSim* can be found in the *LifeSim* Technical Reference Manual [33].

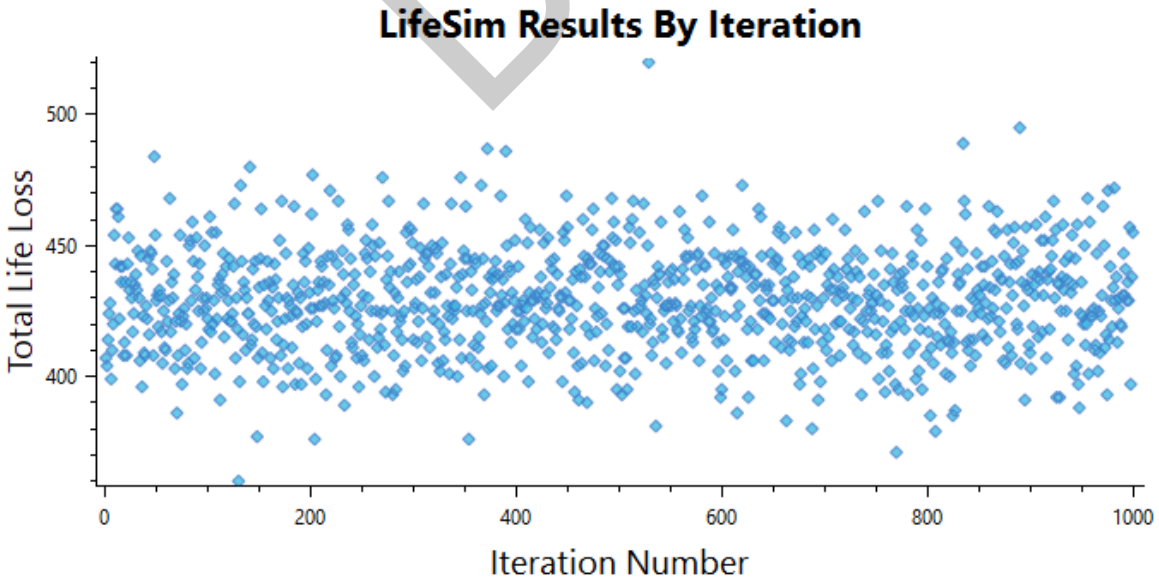


Figure 28 - Life loss results for all iterations from a LifeSim simulation.

Life loss and economic damage results from every simulation are imported from the specified *LifeSim* file. All iterations of the Monte Carlo analysis are included. Once imported, the results are stored in the RMC-TotalRisk study file. If the original *LifeSim* results are updated, they will have to be re-imported. Once imported, a tabular consequence function must be created using the *LifeSim* results. The user creates a tabular consequence function by entering hazard levels in ascending order and selecting the appropriate alternative and time of day for each hazard level as shown in Figure 29 below.

Once a *LifeSim* simulation result has been selected for a given hazard level, multiple distributions are auto-fit to all iterations of the selected data. From the distribution options, by default a truncated normal distribution is selected to represent potential consequences with uncertainty at the specified hazard level. The user can change the distribution and its parameters at any time. The options for defining uncertainty around the *LifeSim* results and the method used to estimate parameters (see **Appendix B** – Summary Statistics) to auto-fit the distribution are defined below:

- **Deterministic**: Method of Moments.
- **Triangular**: Method of Moments.
- **Pert**: Method of Percentiles.
- **Normal**: Method of Moments.
- **Ln-Normal**: Method of Moments.
- **Truncated Normal**: Method of Moments.

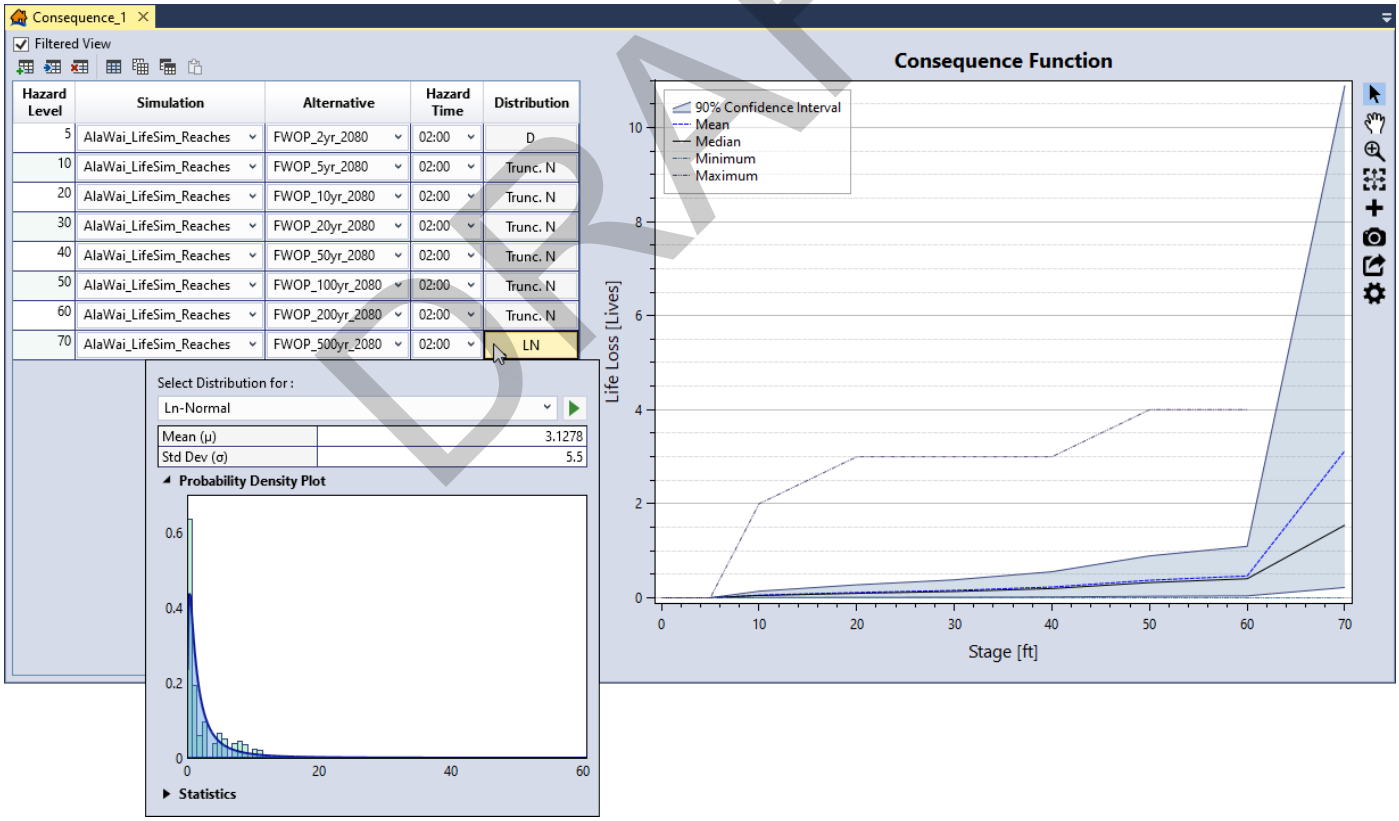


Figure 29 - Example of a tabular function created from LifeSim results.

Composite

There is often a need to combine multiple consequence functions into a single composite function. A composite consequence function can be created by summing across a list of consequences functions as follows:

$$C(x) = \sum_{i=1}^{n} C_i\,(x) \qquad \textit{Equation 47}$$

where $C_i(\cdot)$ is the consequence function i that computes consequences given a hazard level x. This option is particularly useful when damages are estimated separately by types. For example, there could be damages to properties, industry, agriculture, etc. A composite function can be created to aggregate the consequences from all types into a total consequence function.

Alternatively, a composite consequence function can be created by assigning weights (or likelihoods) to a list of consequence functions:

$$C(x) = \sum_{i=1}^{n} \omega_i \cdot C_i\,(x) \qquad \textit{Equation 48}$$

where $C_i(\cdot)$ is the consequence function i that computes consequences given a hazard level x; and ω_i is the weight or likelihood of consequence function i, with $0 \leq \omega_i \leq 1$ and $\sum_{i=1}^{n} \omega_i = 1$.

The user can choose to average the consequence functions based on the user-defined weights, or instead the uncertainty in consequence functions can be modeled as a mixture distribution based on user-defined weights. The mixture distribution approach captures the full uncertainty from all the consequence functions and will result in wider variance and confidence interval widths.

In dam and levee safety, it is common practice to evaluate daytime and nighttime consequences separately, and then assigning a likelihood to each scenario. For example, daytime consequences are typically given a weight of 0.42 (10 hours per day) and nighttime consequences are given a weight of 0.58 (14 hours per day). Figure 30 shows an example of a composite consequence curve for the weighted average of day and night scenarios.

The uncertainty analysis for the composite consequence functions is performed as described in Algorithm 7. This algorithm describes the routine given a single hazard level of interest. It can easily be extended to include a list of hazard levels to evaluate. When simulating risk with full uncertainty, a new composite consequence function is created each Monte Carlo realization, which is then used to derive consequences of failure or non-failure in the risk analysis.

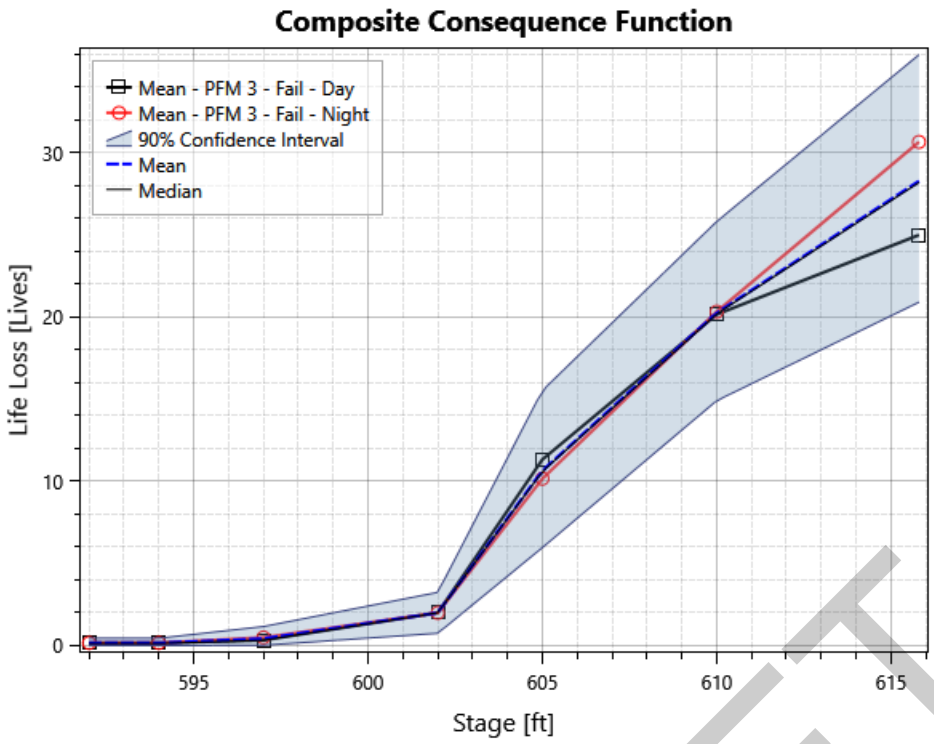


Figure 30 - Example of a composite consequence function using a weighted average for day and night losses.

Algorithm 7 – Composite Consequence Function Uncertainty Analysis

$R \leftarrow$ number of Monte Carlo realizations
$n \leftarrow$ number of consequence functions
$x \leftarrow$ hazard level to evaluate
for $i \leftarrow 1$ to R **do**
  for $j \leftarrow 1$ to n **do**
    Sample uniformly at random with replacement a percentile $r_i \sim \mathrm{U}(0,1)$.
    Sample a random consequence function C_j^* given r_i
    if additive **then**
      $y_i^* \leftarrow y_i^* + C_j^*(x|r_i)$ ▷ Sum the consequences from each function
    else if average **then**
      $y_i^* \leftarrow y_i^* + \omega_j \cdot C_j^*(x|r_i)$ ▷ Average the consequences from each function
    else if mixture **then**
      $F_C \leftarrow F_C + \omega_j$ ▷ Create a cumulative distribution across all functions
      if $r_j \leq F_C$ where $r_j \sim \mathrm{U}(0,1)$ **then** ▷ Randomly sample to determine which function
        $y_i^* \leftarrow C_j^*(x|r_i)$
        break
      end if
    end if
  end for
end for
Estimate confidence intervals and mean consequences for x from $\{y_1^*, \cdots, y_R^*\}$.

Quantitative Risk Analysis

Risk has various definitions and interpretation among different industries, but it is generally understood to describe the probability and severity of an adverse event [2]. Risk can also be defined in terms of an expected value and deviations from the expected value. For example, in the financial industry, the standard deviation is a popular risk measure for stock returns [3]. Flood risk management investment decisions are typically made from a *risk neutral* perspective based on average annual net benefits [4]. As such, flood risk can be formally defined as the expected value of consequences $\mathbb{E}[C]$, which is calculated as:

$$\mathbb{E}[C] = \int_{-\infty}^{\infty} C(x) \cdot f\big(C(x)\big) \cdot dx \qquad \textit{Equation 49}$$

where x is the hazard level (e.g., flood discharge or water level); $C(x)$ determines the consequences, such as property damage or life loss, for the hazard level x; and $f\big(C(x)\big)$ is the probability density function (PDF) of the consequences occurring. The probability distribution of consequences can be defined as a function of hazard, system response, and consequence functions:

$$\mathbb{E}[C] = \int_{-\infty}^{\infty} f_x(x) \cdot F_R(x) \cdot C_R(x) \cdot dx \qquad \textit{Equation 50}$$

where $f_x(x)$ is the probability density function of the hazard (e.g., annual maximum water level or ground acceleration), $F_R(x)$ is the system response function (i.e., probability of failure given hazard level x), and $C_R(x)$ is the consequences given the hazard level x and the system response.

In practice, the risk integral for a dam or levee is often calculated numerically by annualizing and discretizing the hazard. The risk of failure using discrete hazard levels follows from Equation 50 and is defined as:

$$\mathbb{E}[C_F] = \sum_i P(x_i) \cdot P(F|x_i) \cdot C_F(x_i) \qquad \textit{Equation 51}$$

where $P(x_i)$ is the probability of the hazard level x_i; $P(F|x_i)$ is the conditional probability of failure given the hazard level x_i.; and $C_F(x_i)$ is the consequence of failure given the hazard level x_i. Equation 51 is often written semantically to convey the risk equation as:

$$Risk\ of\ Failure\ = P(Hazard) \times P(Failure|Hazard) \times Consequences\ of\ Failure \qquad \textit{Equation 52}$$

where the risk of failure is equal to the probability of the hazard level, $P(Hazard)$, multiplied by the probability of failure given the hazard level, $P(Failure|Hazard)$, multiplied by the consequences of failure at the hazard level, $Consequences\ of\ Failure$.

R-S Reliability Formulation of Risk

This section derives the necessary steps required to go from Equation 49 to Equation 51 using an engineering reliability formulation of risk. The following derivation is taken from [34].

Most engineering risk problems involve two opposing factors: a resistance (or capacity) R at hazard level x, and the load (or demand) S at hazard level x. When the load exceeds the resistance, failure occurs. The word “failure” does not necessarily imply fracture, breach, or collapse of the structure or system. The term here is used in the general sense, meaning that the system fails to meet the demand placed on it. The probability of failure and resulting risk arises due to the uncertainties in the load, resistance, and consequences.

When estimating the risk of failure, the probability of consequences occurring, $f\big(C(x)\big)$, in Equation 49 is assumed to be equivalent to the *probability of failure*, p_F, which is defined as the probability that the resistance is less than or equal to a specific load:

$$p_F = P(R \leq S) \qquad \textit{Equation 53}$$

The complement of the probability of failure, $1 - p_F$, is the *reliability,* or the *probability of non-failure*. The joint PDF of R and S is defined as:

$$p_F = \int \int_{r \leq s} f_{RS}(r,s) \cdot dr \cdot ds \qquad \textit{Equation 54}$$

In RMC-TotalRisk, this double integral is solved by conditioning on S. As such, the joint PDF $f_{RS}(r,s)$ is replaced by its equivalent $f_{R|S}(r|s) \cdot f_S(s)$, where $f_{R|S}(r|s)$ is the conditional PDF of R given the demand $S = s$, and $f_S(s)$ is the marginal PDF of S. The double integral in Equation 54 can then be rewritten as:

$$p_F = \int_{-\infty}^{\infty} \int_{-\infty}^{s} f_{R|S}(r|s) \cdot f_S(s) \cdot dr \cdot ds \qquad \textit{Equation 55}$$

Replacing $\int_{-\infty}^{s} f_{R|S}(r|s) \cdot dr$ with the conditional CDF of R, the double integral reduces to:

$$p_F = \int_{-\infty}^{\infty} F_{R|S}(s|s) \cdot f_S(s) \cdot ds \qquad \textit{Equation 56}$$

which can be seen as an application of the total probability theorem (see **Appendix A**), where the failure event is conditioned on the demand S. As shown in **Appendix A**, the PDF is the derivative of the CDF:

$$f(x) = \frac{dF(x)}{dx} \qquad \textit{Equation 57}$$

Therefore, $f_S(s) \cdot ds$ can be replaced by $dF_S(s)$ so that Equation 56 further reduces to:

$$p_F = \int_{-\infty}^{\infty} F_{R|S}(s) \cdot dF_S(s) \qquad \textit{Equation 58}$$

Now, replacing $f\big(C(x)\big)$ in Equation **49** with $F_{R|S}(s) \cdot dF_S(s)$, and replace the demand s by its equivalent hazard level x gives:

$$\mathbb{E}[C_F] = \int_{-\infty}^{\infty} C_F(x) \cdot F_{R|S}(x) \cdot dF_S(x) \qquad \textit{Equation 59}$$

Finally, by rearranging terms and solving Equation 59 with numerical integration as shown in **Appendix D**, the risk of failure becomes:

$$\mathbb{E}[C_F] = \int_a^b dF_S(x) \cdot F_{R|S}(x) \cdot C_F(x) \approx \sum_i P(x_i) \cdot P(F|x_i) \cdot C_F(x_i) \qquad \textit{Equation 60}$$

Which can be seen as equivalent to Equation **51**.

All inputs in RMC-TotalRisk are modeled as continuous distributions and functions. However, to improve the readability of the risk equations, and since risk has traditionally been computed using numerical integration with discrete hazard bins, the discrete form of the risk equation shown in Equation **51** is used for the remainder of this chapter.

Types of Risk

RMC-TotalRisk estimates five different types of risk: 1) risk of failure; 2) risk of non-failure; 3) total risk; 4) incremental risk of failure; and 5) background risk of non-failure. These various types of risk are consistent with those presented in [7] [11] [35]. In the USACE Dam and Levee Safety Programs, incremental risk is referred to as *dam risk* or *levee risk* and total risk is referred to as *flood risk*.

Each of these risk types have different uses in decision making. For example, incremental risk of failure for a given dam or levee is a primary metric for dam and levee safety decisions. Whereas total risk is used for evaluating the annual net benefits designing a new structure, such as a dam or levee. The following subsections provide the mathematical details for each type of risk computed by RMC-TotalRisk.

Risk of Failure

The risk of failure follows from is defined as:

$$\mathbb{E}[C_F] = \sum_i P(x_i) \cdot P(F|x_i) \cdot C_F(x_i) \qquad \textit{Equation 61}$$

where $P(x_i)$ is the probability of the hazard level x_i; $P(F|x_i)$ is the conditional probability of failure given the hazard level x_i; and $C_F(x_i)$ is the consequence of failure given the hazard level x_i.

Risk of Non-Failure

For infrastructure such as dams and levees, there will often be consequences even if the structure does not fail. For example, during a major flood event, a dam could activate the emergency spillway, preventing the dam from reducing downstream flooding. The risk of non-failure is defined as:

$$\mathbb{E}[C_{NF}] = \sum_i P(x_i) \cdot \{1 - P(F|x_i)\} \cdot C_{NF}(x_i) \qquad \textit{Equation 62}$$

where $\{1 - P(F|x_i)\}$ is the probability of non-failure given the hazard level x_i, which is simply the complement of the probability of failure; and $C_{NF}(x_i)$ is the consequence of non-failure given the hazard level x_i.

The risk of non-failure is important to consider when evaluating potential structural rehabilitation and modification measures for dams and levees. For example, to reduce the risk of dam failure from overtopping, the spillway could be widened or made more efficient, which could increase the risk of non-failure.

Total Risk

The total risk is the sum of the risk of failure and non-failure and is defined as follows:

$$\mathbb{E}[C_T] = \sum_i P(x_i) \cdot [P(F|x_i) \cdot C_F(x_i) + \{1 - P(F|x_i)\} \cdot C_{NF}(x_i)] \quad \textit{Equation 63}$$

Total risk can also be computed by summing the risk of failure and risk of non-failure:

$$\mathbb{E}[C_T] = \mathbb{E}[C_F] + \mathbb{E}[C_{NF}] \quad \textit{Equation 64}$$

In some USACE publications, total risk has been referred to as *residual risk*, i.e., the risk that remains [7] [11] [36]. The total risk is often used to evaluate risk reduction alternatives for new FRM structures and systems. In addition, risk reduction alternatives should *do no harm* by not increasing total risk [11]. For example, to reduce the risk of dam failure from overtopping, the spillway could be widened or made more efficient, which could increase the risk of non-failure. The spillway alternative should not increase the risk of non-failure more than it reduces the risk of failure, i.e., the total risk should decrease and not increase.

Incremental Risk

A common practice when evaluating rehabilitation or modification of existing infrastructure for safety is to estimate risk using incremental consequences [11] [35] [37]. Incremental consequences are the incremental losses or damages that failure might inflict over and above any damage which might occur for the same event assuming the structure does not fail. In USACE safety programs, incremental risk is the primary risk measure used for decision making and portfolio prioritization because it represents the risk that can be reduced by modifications within existing authorizations from Congress. The incremental risk equation is:

$$\mathbb{E}[C_\Delta] = \sum_i P(x_i) \cdot P(F|x_i) \cdot C_\Delta(x_i) \quad \textit{Equation 65}$$

where $C_\Delta(x_i) = C_F(x_i) - C_{NF}(x_i)$. Incremental risk is sometimes referred to as *excess risk*. In certain simulation scenarios, there is the possibility that the simulated consequences of non-failure are greater than consequences of failure, leading to negative incremental consequences. In RMC-TotalRisk, negative incremental consequences (and thus negative incremental risk) are not permitted, so negative incremental consequences are set to zero if they occur during a risk simulation. RMC-TotalRisk only accounts for consequences of failure *over and above* or *in excess of* consequences of non-failure.

Background Risk

Background risk is defined as the risk of the structure assuming that it has no structural flaws or vulnerabilities. In other words, if all the structural vulnerabilities were eliminated, the remaining risk would be the background risk from natural hazards. The background risk equation can be derived from the total risk equation:

$$\mathbb{E}[C_B] = \sum_i P(x_i) \cdot [P(F|x_i) \cdot C_F(x_i) + \{1 - P(F|x_i)\} \cdot C_{NF}(x_i)] \quad \textit{Equation 66}$$

Because there are assumed to be no structural vulnerabilities, the probability of failure is set equal to zero, $P(F|x_i) = 0$, for all potential failure modes. This then leads to the following equation for background risk:

$$\mathbb{E}[C_B] = \sum_i P(x_i) \cdot [0 \cdot C_F(x_i) + 1 \cdot C_{NF}(x_i)] \quad \textit{Equation 67}$$

which further reduces to:

$$\mathbb{E}[C_B] = \sum_i P(x_i) \cdot C_{NF}(x_i) \qquad \textit{Equation 68}$$

In some USACE publications, background risk has been referred to as *non-breach* risk [11]. To avoid confusion with the risk of non-failure, the term *background risk* is used in the RMC-TotalRisk software.

Incremental risk and background risk are risk measures commonly used for prioritizing safety investment decisions. However, it is important to note that these risk types are not representative of real events that could occur at the structure being evaluated. Risk of failure, non-failure, and total risk represent the risks for actual events experienced by the people and property at risk.

For a single system component, Total risk can also be computed by summing the incremental risk and background risk:

$$\mathbb{E}[C_T] = \mathbb{E}[C_\Delta] + \mathbb{E}[C_B] \qquad \textit{Equation 69}$$

In RMC-TotalRisk, a system can have multiple components, each with multiple failure modes where each component also has a separate non-failure mode. As such, there is a potential for some embedded correlation between incremental consequences and non-failure consequences across system components. Therefore, for a system with multiple components, depending on the selected joint consequence rule, total risk will be greater than or equal to incremental plus background risk. More details on system risk are provided in the next section.

System Risk

The previous section described the types of risk that can be evaluated for a single system component involving a single failure mode. This section will demonstrate how those risk types can be generalized for complex multicomponent systems.

Most engineering risk and reliability problems involve complex systems composed of multiple components each with multiple failure modes. For example, a levee system might have several independent levee segments, each with several potential failure modes. Most physical engineering systems that are composed of multiple components can be classified as series-connected or parallel-connected systems, or combinations thereof [9] [38].

In a *series system*, if any component fails, the entire system fails. In other words, a series system is a nonredundant system, also known as a “weakest link” system. Figure 31 provides a schematic of a series system, which resembles links in a chain; if one link fails, the whole chain fails. Many engineering systems are in series, such as highways, bridges, pipelines, levees, dams, etc.

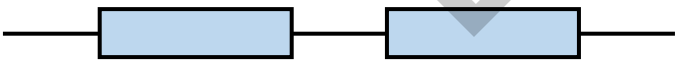


Figure 31 - Schematic of a series system.

Parallel systems are redundant systems that only fail when all components in the system fail. Figure 32 provides a schematic of a parallel system.

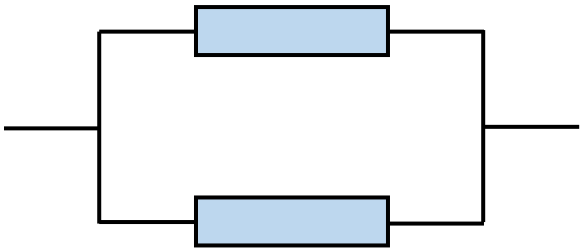


Figure 32 - Schematic of a parallel system.

Engineering systems may be composed of a combination of series-connected and parallel-connected components as shown in Figure 33.

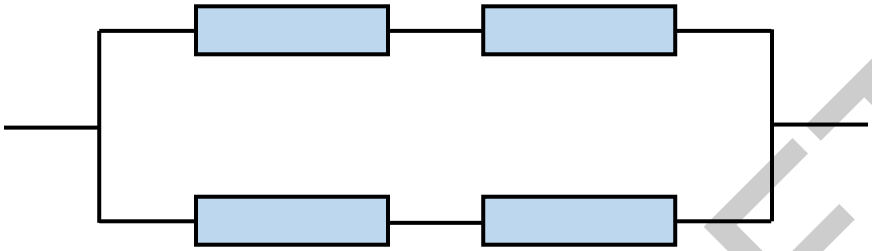


Figure 33 - Schematic of a combination system.

The probability of failure of a series system is the union of all the failure states of the components:

$$P(F|x) = P\left(\bigcup_i F_i|x\right) \qquad \textit{Equation 70}$$

where $F_i|x$ is failure for failure mode i given the hazard level x. The probability of failure of a parallel system is the intersection (or joint probability) of all the failure states:

$$P(F|x) = P\left(\bigcap_j F_j|x\right) \qquad \textit{Equation 71}$$

The probability of failure for a combination system is:

$$P(F|x) = P\left(\bigcup_i \bigcap_j F_{i,j}|x\right) \qquad \textit{Equation 72}$$

If all parallel-connected i failures are independent, and all series-connected j failures are mutually exclusive, the probability of failure is:

$$P(F|x) = \sum_j \prod_i P\left(F_{i,j}|x\right) \qquad \textit{Equation 73}$$

This *sum-product* formulation is the same as Equation 40, which is used to evaluate event trees.

In RMC-TotalRisk, multiple failure modes are treated as a series system, consistent with [9] and [39]. Parallel modes of failure can be modeled with event trees, as shown in **System Response Functions** chapter. The following subsections describe how multiple failure modes and multiple system components are incorporated into the RMC-TotalRisk system risk analysis.

Risk Diagram

A risk analysis in RMC-TotalRisk is defined through a diagram as shown in Figure 34 below. The diagram provides an intuitive way to create and connect the various components of the modeled system. Figure 34 shows a single system component for a dam safety risk analysis. There is a non-failure mode, shown at the top of the diagram with the purple line, that connects the hazard function to the non-failure consequences, without any system response. For many dams, there will often be consequences even if the structure does not fail. For example, during a major flood event, a dam could activate the emergency spillway, preventing the dam from reducing downstream flooding. The non-failure mode is used to model the risk of non-failure. There are two failure modes: 1) A spillway erosion failure mode, labelled PFM 1, shown in the center of the diagram connects the hazard function at Dam A to the PFM 1 response function and consequence function; and 2) A concentrated leak erosion failure mode, labelled PFM 2, shown in the bottom of the diagram with the same respective connections.

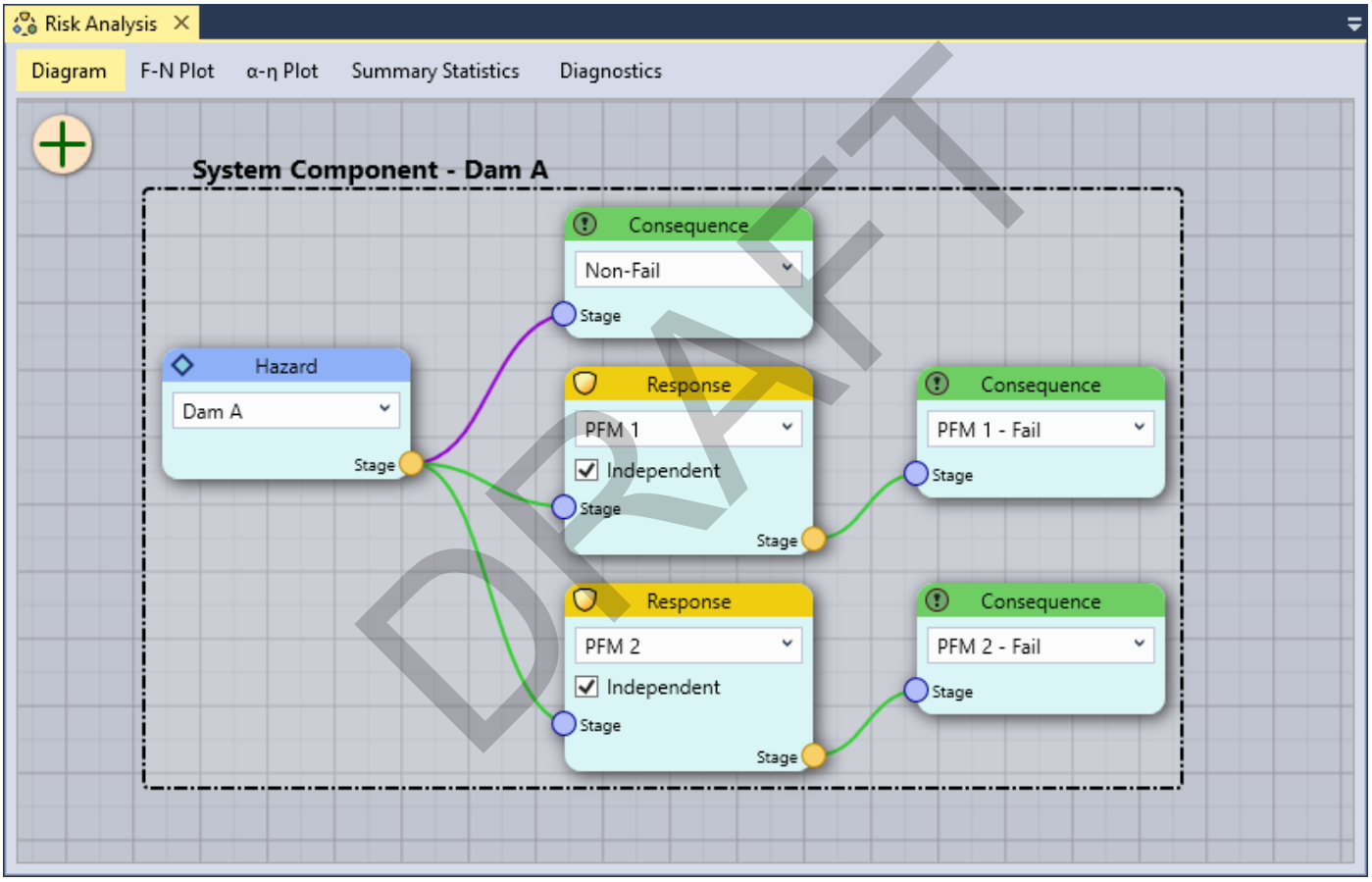


Figure 34 – RMC-TotalRisk risk diagram.

In the risk diagram, the input functions are connected from left to right, and linked together based on the hazard type (e.g., stage, flow, etc.) and units (e.g., ft, cfs, etc.) of each function. There cannot be any circular or redundant connections. As such, the risk diagram is a type of directed acyclic graph (DAG).

The system components are identified and labeled by the selected hazard function. The failure modes within a component are identified and labeled by the selected response functions. RMC-TotalRisk permits an unlimited number of failure modes per component, depending on the selected failure mode method. However, a single system is limited to 20 components due to virtual memory and computer runtime limitations. For example, the system risk of a watershed comprising up to 20 dams, each with 20 failure modes, can be assessed. More details on these constraints are provided in the following subsections.

Multiple Failure Modes

A failure mode for a system component is defined as some unique combination of events whose joint occurrence would cause the system to fail [39]. The conditional probability of failure given a hazardous event, $P(F|x)$, is often a function of multiple failure modes, some of which could have a positive dependency and others could have a negative dependency. For example, internal erosion driven failure modes for a dam or levee typically have a positive dependency [39]. In other words, the dam or levee can be in the process of failure from multiple internal erosion failure modes simultaneously. On the other hand, if the dam or levee is failing from overtopping, it may be less likely, or not even possible, that it will also fail from internal erosion failure modes.

In RMC-TotalRisk, multiple failure modes are modeled as a system in series because any one of them will cause failure of the system component [39]. For this reason, the overall conditional probability of failure $P(F|x)$ must be computed using the *probability of union* (see **Appendix A**). For example, if there are two failure modes, the probability of union is:

$$P(F|x) = P\left(\bigcup_i F_i|x\right) \qquad \textit{Equation 74}$$

$$P(F|x) = P(F_1|x) + P(F_2|x) - P(F_1|x \cap F_2|x) \qquad \textit{Equation 75}$$

where $P(F_1|x)$ is the conditional probability of failure for failure mode 1; $P(F_2|x)$ is the conditional probability of failure for failure mode 2; and $P(F_1|x \cap F_2|x)$ is the joint probability of failure mode 1 and 2 both occurring at the same hazard level.

In practice, there is rarely enough data to precisely estimate the dependence between failure modes. Instead, the probability of union between failure modes can be bounded following the *Unimodal Bounds Theorem* [9] [38] [39]. These bounds are sometimes referred to as *Fréchet inequalities* or *risk bounds* [40]. For positively dependent failure modes, the true conditional probability of failure $P(F|x)$ will be within the range:

$$\max_i\{P(F_i|x)\} \leq P(F|x) \leq 1 - \prod_{i=1}^{k}[1 - P(F_i|x)] \qquad \textit{Equation 76}$$

If the failure modes are perfectly positively dependent, the true probability of failure will equal the most probable failure mode, $\max_i\{P(F_i|x)\}$. On the other hand, if the failure modes are statistically independent, the true probability of failure will equal the probability of union following *De Morgan's rule* and the rule of subtraction for k failure modes. If the failure modes are negatively dependent, the true probability of failure will be within the range:

$$1 - \prod_{i=1}^{k}[1 - P(F_i|x)] \leq P(F|x) \leq \min\left(1, \sum_{i=1}^{k} P(F_i|x)\right) \qquad \textit{Equation 77}$$

where $\sum_{i=1}^{k} P(F_i|x)$ is the probability of union for mutually exclusive failure modes. In practice, independence is assumed for positively dependent failure modes to err on the side of caution. Likewise, perfect negative dependency is assumed for negatively dependent failure modes.

The unimodal bounds provided in Equation 76 and Equation 77 are sometimes too wide for use in quantitative risk analysis. In addition, these bounds are only of practical use if each failure mode has the same consequence function [41]. Consequences of failure can vary widely depending on the failure mechanism. The unimodal bound approach combines the probability of failure for each mode up front before evaluating the consequences for each failure mode. However, in most quantitative risk analyses, it is often desirable to keep track of the consequences for each individual failure mode so that risk reduction alternatives can be evaluated.

When each failure mode has different levels of consequences, more robust techniques must be used. RMC-TotalRisk provides three computation methods for assessing multiple failure modes: 1) the common cause adjustment, 2) competing failure modes, and 3) joint failure modes.

Common Cause Adjustment

The Common Cause Adjustment (CCA) is a method that was first applied to dam safety risk analyses [42] [41]. The method was originally intended for failure modes that are not mutually exclusive and that can occur simultaneously at multiple sections of a dam due to a single or common cause initiating event [42]. The failure modes are assumed to be positively correlated and conditioned on the same loading event; hence they have a *common cause*. The upper unimodal bound for positively correlated failure modes is *De Morgan's rule*. However, that equation does not provide a direct method for computing the risk of failure when each failure mode has unique consequences. To improve upon the unimodal bounds, the CCA method reapportions the conditional probabilities of failure of each failure mode such that each can be modelled as being mutually exclusive. First, the probability of union assuming independence (*De Morgan's rule*) is set equal to the probability of union assuming mutual exclusivity by multiplying by a constant c:

$$1 - \prod_{i=1}^{k}[1 - P(F_i|x)] = \; c \cdot \sum_{i=1}^{k} P(F_i|x) \qquad \textit{Equation 78}$$

Rearranging, the constant (or adjustment factor) is calculated as:

$$c = \frac{1 - \prod_{i=1}^{k}[1 - P(F_i|x)]}{\sum_{i=1}^{k} P(F_i|x)} \qquad \textit{Equation 79}$$

where $0 \leq c \leq 1$. Now, to compute the combined probability of failure each marginal conditional probability is multiplied by the constant:

$$P(F|x) = \sum_{i=1}^{k} c \cdot P(F_i|x) \qquad \textit{Equation 80}$$

Then, the risk of failure with k potential failure modes is as follows:

$$\mathbb{E}[C_F] = \sum_{i} P(x_i) \cdot \sum_{j=1}^{k} c \cdot P\left(F_j|x_i\right) \cdot C_{F_j}(x_i) \qquad \textit{Equation 81}$$

As an illustrative example, consider three failure modes A, B, and C with probabilities of failure 0.25, 0.35, and 0.50, respectively. A Venn diagram of these failure modes is shown in Figure 35 below. These failure modes are not mutually exclusive and therefore overlap in the diagram.

The probability of union assuming independence is:

$$P(A \cup B \cup C) = \; 1 - \; [1 - P(A)] \cdot [1 - P(B)] \cdot [1 - P(C)] \qquad \textit{Equation 82}$$

$$P(A \cup B \cup C) = \; 1 - \; [1 - 0.25] \cdot [1 - 0.35] \cdot [1 - 0.50] = 0.75625 \qquad \textit{Equation 83}$$

The probability of union assuming mutual exclusivity is:

$$P(A \cup B \cup C) = \; P(A) + P(B) + P(C) \qquad \textit{Equation 84}$$

$$P(A \cup B \cup C) = 0.25 + 0.35 + 0.50 = 1.10 \qquad \textit{Equation 85}$$

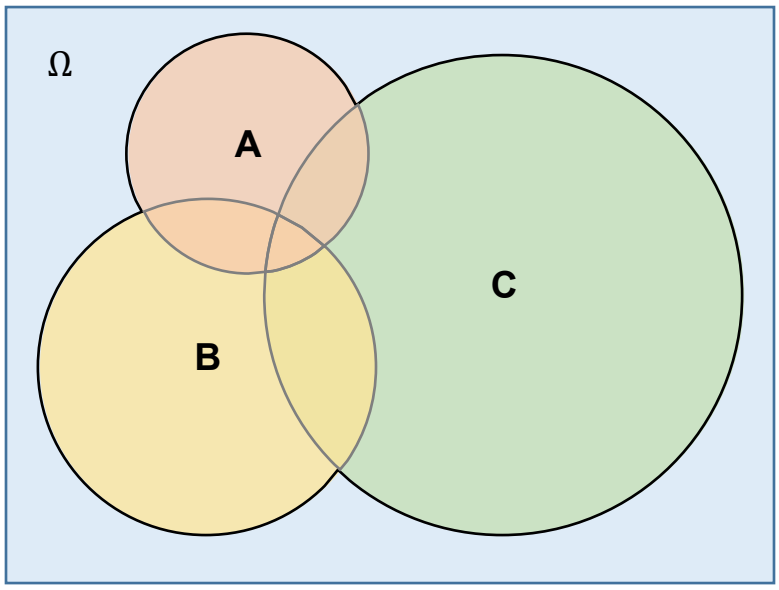


Figure 35 - Venn diagram for the union of three failure modes.

The CCA adjustment factor is:

$$c = \frac{0.75625}{1.10} = 0.6875$$ *Equation 86*

Then the adjusted conditional probabilities are calculated as:

$$P(A \cup B \cup C) = c \cdot P(A) + c \cdot P(B) + c \cdot P(C)$$ *Equation 87*

$$P(A \cup B \cup C) = 0.171875 + 0.240625 + 0.343750 = 0.75625$$ *Equation 88*

These adjusted conditional probabilities of failure are visualized in Figure 36. The probabilities are reduced proportionally so that they no longer intersect, while still summing to the same probability of union.

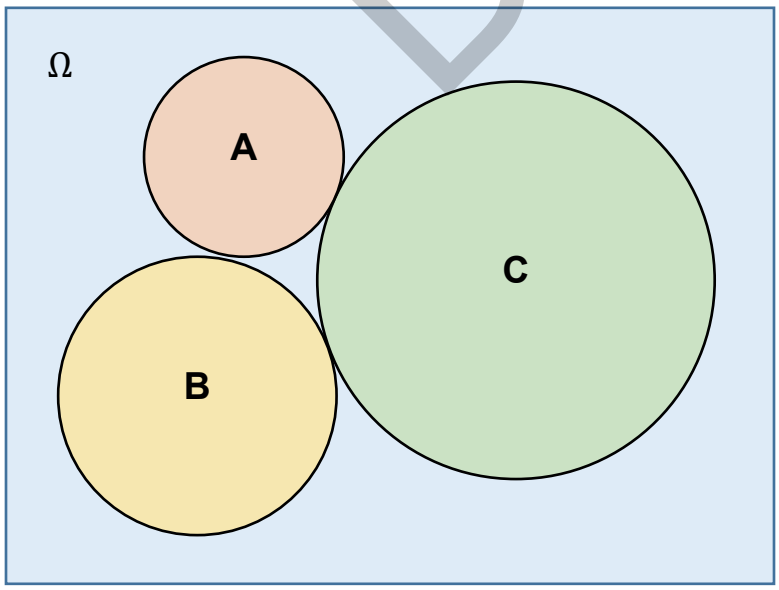


Figure 36 - Venn diagram for the union of three failure modes after applying the Common Cause Adjustment.

The CCA was originally intended for positively correlated or independent failure modes. However, there are situations where failure modes can be negatively dependent, which will lead to a higher combined probability of failure for the system [39] [43]. RMC-TotalRisk employs a generalized version of the CCA that can work with any dependency. The generalized adjustment constant is:

$$c = \frac{P(\bigcup_{i=1}^{n} F_i|x)}{\sum_{i=1}^{n} P(F_i|x)} \qquad \textit{Equation 89}$$

The probability of union in the numerator is calculated using the inclusion-exclusion principle as discussed in the **Joint Failure Modes** section and in **Appendix A**.

The CCA is the default failure mode method in RMC-TotalRisk. This method is flexible, it can incorporate dependencies between failure modes, it allows for an unlimited number of failure modes, and it is very computationally efficient. Although, it is important to note that the CCA method does not have a strong theoretical basis.

The key assumption of the CCA method is that failure modes are not mutually exclusive, but instead dependent and likely to occur simultaneously. However, rather than explicitly accounting for the joint occurrence of failures, the CCA method reapportions the conditional probabilities of failure such that each failure mode can be treated as mutually exclusive at the point of first failure. This eliminates the possibility of joint failures and the possibility for joint consequences of failure.

Because there is not a strong theoretical basis, the CCA method has the potential to lead to undesirable behavior and unrealistic failure probabilities for individual failure modes. For example, there are two special cases that can arise: *physical dominance* and *freezing* [42] [41] [44]. An example of *dominance* is where one failure mode develops more rapidly at a lower hazard level than the others, making the others less likely. For example, a dam could breach due to internal erosion at a normal pool elevation, which reduces the reservoir level, and thereby reduces the chance of simultaneous failures from other failure modes. *Freezing* refers to an approach where the CCA adjusted probabilities of failure are held constant for all hazard levels above the first hazard level for which at least one unadjusted failure probability becomes 1.0 [44]. Both *dominance* and *freezing* can be more formally addressed using the competing failure modes method described in the next section.

Competing Failure Modes

A competing failure analysis represents a combination of two or more failure modes that are “competing” to the end of life of a series system. The following three key assumptions are required for competing failure mode analysis:

1. Each failure mode proceeds independently of every other one until failure occurs. At the point of first failure, each failure mode is mutually exclusive from one another; i.e., there cannot be joint failures.

2. The system component fails when the first from all the competing failure modes reaches a failure state. The first failure precludes the other failure modes from occurring.

3. Each of the k failure modes has a known failure distribution, $F(\cdot)$, which is monotonically increasing with increasing hazard levels.

The competing failure mode approach can be thought of as a race to see which failure mode will fail first. The failure modes have no interaction or awareness of the state of the other failure modes and are unaware if one has already initiated or not. Each failure mode progresses independently from the others until the first failure causes the system component to fail. With this approach, two or more failure modes can never reach a failure state simultaneously, i.e., there can be no ties in the race, there must always be a single winner.

In public health research, diseases and illnesses are treated as competing risks, where each illness is competing to end the life of the patient. For example, a patient can die from breast cancer or from a stroke but cannot die from both.

In the context of dam and levee safety, an internal erosion failure mode can be thought of as competing with an overtopping failure mode. A levee can only fail from either internal erosion or overtopping, not both. Since these failure modes are competing, a levee failure prior to overtopping due to internal erosion can significantly reduce the probability of failure and consequences of an overtopping failure mode if the prior to overtopping failure fills the leveed area before overtopping occurs.

The formula for a competing risk model is typically written in terms of the survival function, $S(\cdot)$. The survival function for a single failure mode i is the complement of the CDF:

$$S_i(x) = 1 - F_i(x) = 1 - P(F_i|x) \quad \textit{Equation 90}$$

Assuming statistical independence between failure modes, the survival function for the system component is:

$$S_c(x) = \prod_{i=1}^{k} S_i(x) \quad \textit{Equation 91}$$

The CDF of the failure distribution for the system component follows from the union of each failure mode:

$$F_c(x) = 1 - S_c(x) = 1 - \prod_{i=1}^{k} [1 - P(F_i|x)] \quad \textit{Equation 92}$$

The cumulative incidence function (CIF) is the cumulative probability of failure from a specific failure mode. In the literature, the CIF is also referred to as the sub-distribution function because it is not a true probability distribution. The CIF for a single failure mode is computed as follows:

$$CIF_i(x) = \int_{-\infty}^{x} h_i(t) \cdot S_c(t) \cdot dt \quad \textit{Equation 93}$$

where $h_i(\cdot)$ is the hazard function for failure mode i:

$$h_i(x) = \frac{f_i(x)}{1 - F_i(x)} = \frac{f_i(x)}{S_i(x)} \quad \textit{Equation 94}$$

The CDF of the failure distribution for the system component can also be computed by summing the CIFs:

$$F_c(x) = \sum_{i=1}^{k} \int_{-\infty}^{x} h_i(t) \cdot S_c(t) \cdot dt \quad \textit{Equation 95}$$

$$F_c(x) = \sum_{i=1}^{k} CIF_i(x) \quad \textit{Equation 96}$$

In RMC-TotalRisk, Equation 93 is pre-processed for each failure mode with numerical integration using $n = 200$ partitioned hazard levels as follows:

$$CIF_i(x) = \sum_{j=2}^{n} \frac{\{P(F_i|x_j) - P(F_i|x_{j-1})\}}{\overline{S_t(x)}} \cdot \overline{S_c(x)} \quad \textit{Equation 97}$$

where $\overline{S(\cdot)}$ is the mean survival function evaluated for the hazard level x. This is done as a pre-processing step before the adaptive integration routine used to solve the overall risk integral. Please see **Appendix D** for more details on numerical integration.

In RMC-TotalRisk, the competing failure mode method allows for an unlimited number of failure modes, and it is very computationally efficient. The competing failure mode method is similar to the CCA method. Each failure mode is treated as statistically independent from one another until the point of first failure, at which point the failure modes are treated as mutually exclusive. Although, the competing failure mode approach, which is typically referred to as *competing risks*, has a stronger theoretical basis in statistics and risk analysis.

As mentioned before, the competing failure mode method can properly account for *physical dominance* and *freezing*. The failure probability for a single failure mode is adjusted to account for the survival probabilities of the other failure modes (Equation 93). Therefore, the CIF inherently accounts for dominance at lower hazard levels. In addition, the CIF behaves like a proper CDF where the nonexceedance probability is monotonically increasing with increasing hazard levels, and the sum of the CIFs will never exceed 1.0. As such, there is no need for freezing adjustment with this method.

However, the superior theoretical approach of the competing failures method requires more rigid input assumptions: All potential failure modes are treated as statistically independent, and each system response function must be monotonically increasing with increasing hazard levels.

Joint Failure Modes

Unlike the previous two methods, the joint failure modes method directly allows for dependency between failure modes and allows for multiple (joint) failures during the same hazard event. The only assumption required for this analysis is that a rule must be applied to account for the consequences of a joint failure.

Returning to the previous example, consider three failure modes A, B, and C with probabilities of failure 0.25, 0.35, and 0.50, respectively. Because these failure modes are not mutually exclusive and sum to greater than 1, an event tree can be constructed to show all possible failure and non-failure pathways, as shown in Figure 37 below.

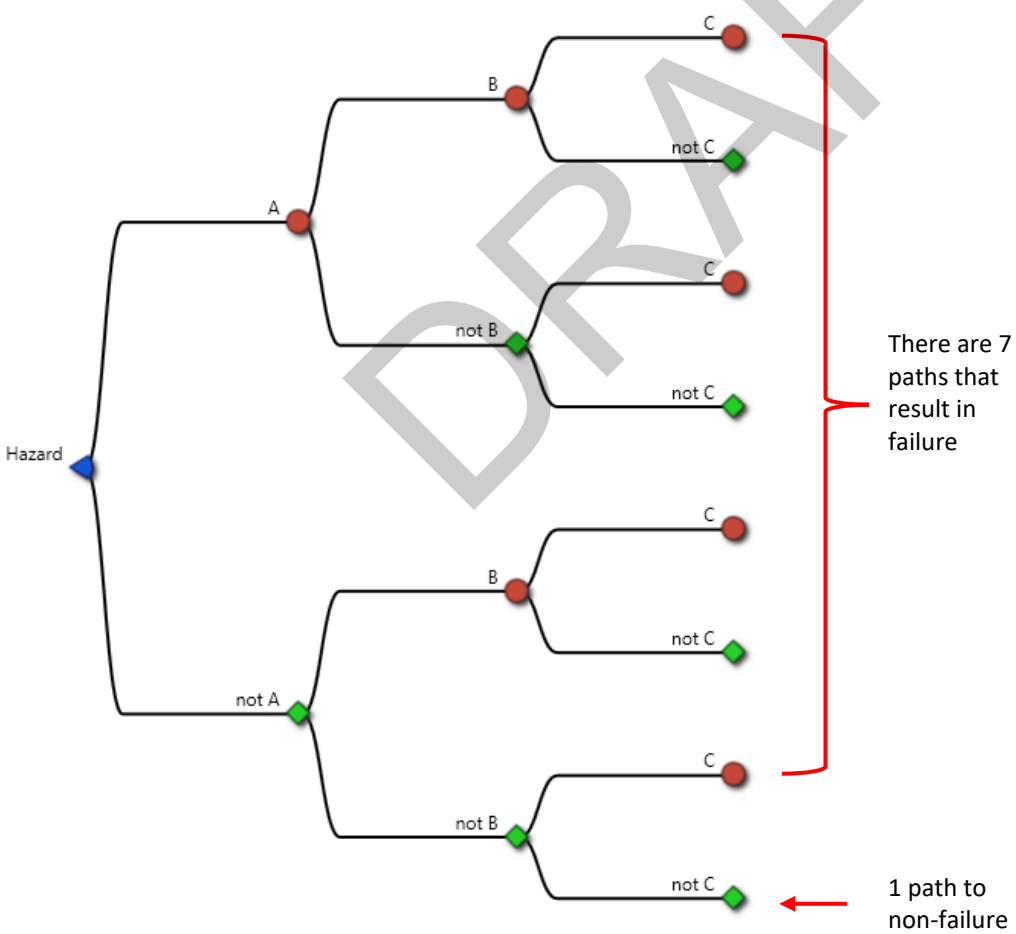


Figure 37 – Event tree diagram for three failure modes showing all possible pathways.

When there are multiple failure modes, the number of possible ways the system can fail is $2^n - 1$. For example, as seen in Figure 37, if there are 3 failure modes, there are $2^3 - 1 = 7$ pathways that result in system failure, and one pathway to non-failure. As the number of failure modes and system components increase, the number of ways the system can fail increases exponentially (see **Appendix A** for more details). Considering this, in RMC-TotalRisk, if the joint failure mode option is selected, the maximum number of failure modes allowable for a single system component is 20.

The event tree in Figure 37 can also be decomposed into a table as shown in Table 5 below. The first column provides a binary indicator for each unique failure combination. The second column shows the corresponding symbolic representation of each unique pathway in the tree. The third column shows the pathway probability assuming the failure modes are statistically independent. If the indicator in the first column is 1, then multiply by the respective marginal probability. If the indicator is 0 then multiply by the complement of the marginal probability.

Table 5 - Table for three failure modes showing all possible pathways and joint probabilities.

Combinations	Pathways	Pathway Probability
1 0 0	$P(A \cap \bar{B} \cap \bar{C})$	$= 0.25 \cdot (1 - 0.35) \cdot (1 - 0.50) = 0.08125$
0 1 0	$P(\bar{A} \cap B \cap \bar{C})$	$= (1 - 0.25) \cdot 0.35 \cdot (1 - 0.50) = 0.13125$
0 0 1	$P(\bar{A} \cap \bar{B} \cap C)$	$= (1 - 0.25) \cdot (1 - 0.35) \cdot 0.50 = 0.24375$
1 1 0	$P(A \cap B \cap \bar{C})$	$= 0.25 \cdot 0.35 \cdot (1 - 0.50) = 0.04375$
1 0 1	$P(A \cap \bar{B} \cap C)$	$= 0.25 \cdot (1 - 0.35) \cdot 0.50 = 0.08125$
0 1 1	$P(\bar{A} \cap B \cap C)$	$= (1 - 0.25) \cdot 0.35 \cdot 0.50 = 0.13125$
1 1 1	$P(A \cap B \cap C)$	$= 0.25 \cdot 0.35 \cdot 0.50 = 0.04375$
0 0 0	$P(\bar{A} \cap \bar{B} \cap \bar{C})$	$= (1 - 0.25) \cdot (1 - 0.35) \cdot (1 - 0.50) = 0.24375$

If the failure modes have dependency, computing the pathway probabilities is not as simple as the method provided above in Table 5. To correctly account for dependency, RMC-TotalRisk uses the *inclusion-exclusion principle* (see **Appendix A**), which generalizes the formula for computing the probability of union for many events with dependency as follows:

$$P(F|x) = P\left(\bigcup_{i=1}^{n} F_i|x\right) = \sum_{k=1}^{n} (-1)^{k+1} \cdot \left(\sum_{1 \le i_1 < \cdots < i_k \le n} P\left(F_{i_1}|x \cap \cdots \cap F_{i_k}|x\right)\right) \qquad \textit{Equation 98}$$

This equation can be more clearly understood by visualizing the case of three events A, B, and C with a Venn diagram as shown in Figure 38. The probability union for the three events following the inclusion-exclusion principle is given by:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \qquad \textit{Equation 99}$$

$$P(A \cup B \cup C) = 0.25 + 0.35 + 0.50 - (0.25 \cdot 0.35) - (0.25 \cdot 0.50) - (0.35 \cdot 0.50) + (0.25 \cdot 0.35 \cdot 0.50) = 0.75625 \qquad \textit{Equation 100}$$

This answer is the same as the probability of union using *De Morgan's* rule and shown above in Equation 88.

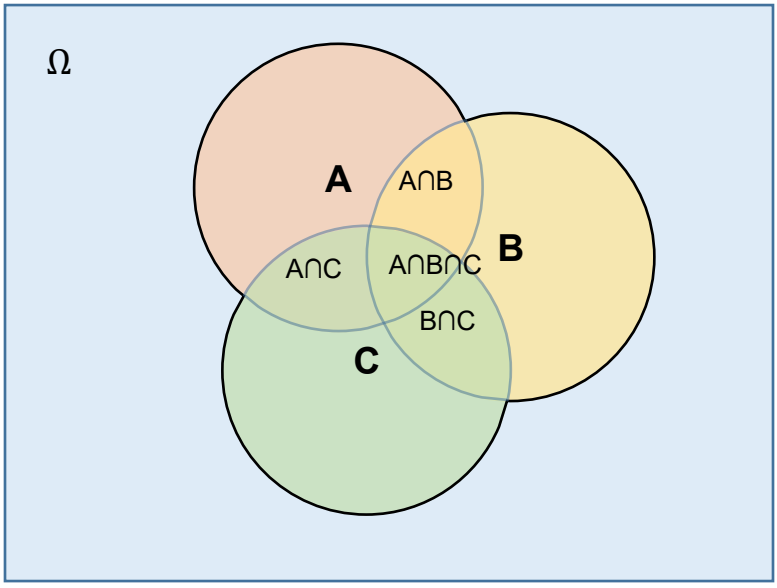


Figure 38 - Venn diagram for the union of three events.

The inclusion-exclusion principle can be generalized into the following recursive steps:

1. Include the marginal probability of each event, $P(A) + P(B) + P(C)$
2. Exclude the pairwise joint probability, $-P(A \cap B) - P(A \cap C) - P(B \cap C)$
3. Include the triple-wise joint probability, $+P(A \cap B \cap C)$
4. Continue until the n-tuple-wise joint probability is included (if n is odd) or excluded (if n is even)

This formula calculates the total probability of union with consideration of dependency. However, this formula does not provide the individual pathway probabilities. To get the correct pathway probabilities, the inclusion-exclusion method must be reversed as follows:

1. Compute the triple-wise joint probability, $P(A \cap B \cap C)$
2. Compute each pairwise joint probability and exclude the triple-wise joint probability, e.g., $P(A \cap B \cap \bar{C}) = P(A \cap B) - P(A \cap B \cap C)$
3. Compute each marginal probability, exclude the associated pairwise joint probabilities, and include the triple-wise joint probability, e.g., $P(A \cap \bar{B} \cap \bar{C}) = P(A) - P(A \cap B) - P(A \cap C) + P(A \cap B \cap C)$

The first column in Table 6 below shows the marginal, pairwise, and triple-wise combinations. The second column provides the probability of each event assuming independence. The probability of each pathway can be computed by reversing the inclusion-exclusion principle as shown in the fourth column. These results are identical to the results shown in Table 5. The key advantage of this approach is that dependencies between failure modes can be directly accounted for using the Multivariate Normal distribution as described in **Appendix F**.

Table 6 - Table for three failure modes showing all possible pathways and joint probabilities using the inclusion-exclusion principle.

Marginal Events	Probability	Pathways	Pathway Probability
$P(A)$	$= 0.25$	$P(A \cap \bar{B} \cap \bar{C})$	$= 0.25 - 0.0875 - 0.1250 + 0.04375 = 0.08125$
$P(B)$	$= 0.35$	$P(\bar{A} \cap B \cap \bar{C})$	$= 0.35 - 0.0875 - 0.1750 + 0.04375 = 0.13125$
$P(C)$	$= 0.50$	$P(\bar{A} \cap \bar{B} \cap C)$	$= 0.50 - 0.1250 - 0.1750 + 0.04375 = 0.24375$
$P(A \cap B)$	$= 0.25 \cdot 0.35 = 0.0875$	$P(A \cap B \cap \bar{C})$	$= 0.0875 - 0.04375 = 0.04375$
$P(A \cap C)$	$= 0.25 \cdot 0.50 = 0.1250$	$P(A \cap \bar{B} \cap C)$	$= 0.1250 - 0.04375 = 0.08125$
$P(B \cap C)$	$= 0.35 \cdot 0.50 = 0.1750$	$P(\bar{A} \cap B \cap C)$	$= 0.1750 - 0.04375 = 0.13125$
$P(A \cap B \cap C)$	$= 0.25 \cdot 0.35 \cdot 0.50 = 0.04375$	$P(A \cap B \cap C)$	$= 0.25 \cdot 0.35 \cdot 0.50 = 0.04375$
$P(A \cup B \cup C)$	$= 0.75625$	$P(\bar{A} \cap \bar{B} \cap \bar{C})$	$= 1 - 0.75625 = 0.24375$

A rule must be assumed to account for the joint consequences of failure. RMC-TotalRisk includes four joint consequence options:

- **Additive**: the consequences for joint failures are calculated as the sum of the consequences for each failure mode that occurs
- **Average**: the consequences for joint failures are calculated as the average of the consequences for each failure mode that occurs.
- **Maximum**: the consequences for joint failures are calculated as the maximum of the consequences over the failure modes that occurs.
- **Minimum**: the consequences for joint failures are calculated as the minimum of the consequences over the failure modes that occur.

In the context of dam and levee safety, additive consequences can be useful when each failure mode has a different inundation area where the consequences for the joint failure are best represented by adding the consequences for each inundation area. Average consequences can be used if the inundation areas are partially overlapping. If each failure mode has practically the same inundation area, then the maximum rule can be used so that consequences are not overestimated. The minimum rule can be used as a sensitivity or as a lower bound for risk.

Following the previous example, failure mode A will result in 10 lives lost, failure mode B in 3 lives lost, and failure mode C in 2 lives lost. Using the additive consequence rule, the expected consequences of failure are computed as shown in Table 7.

Table 7 - Table for three failure modes showing all possible pathways and the expected consequences of failure.

Combinations	Pathways	Pathway Probability	Consequences	Expected
1 0 0	$P(A \cap \bar{B} \cap \bar{C})$	0.08125	$= 10$	$= 0.08125 \cdot 10 = 0.81250$
0 1 0	$P(\bar{A} \cap B \cap \bar{C})$	0.13125	$= 3$	$= 0.13125 \cdot 3 = 0.39375$
0 0 1	$P(\bar{A} \cap \bar{B} \cap C)$	0.24375	$= 2$	$= 0.24375 \cdot 2 = 0.48750$
1 1 0	$P(A \cap B \cap \bar{C})$	0.04375	$= 10 + 3 = 13$	$= 0.04375 \cdot 13 = 0.56875$
1 0 1	$P(A \cap \bar{B} \cap C)$	0.08125	$= 10 + 2 = 12$	$= 0.08125 \cdot 12 = 0.97500$
0 1 1	$P(\bar{A} \cap B \cap C)$	0.13125	$= 3 + 2 = 5$	$= 0.13125 \cdot 5 = 0.65625$
1 1 1	$P(A \cap B \cap C)$	0.04375	$= 10 + 3 + 2 = 15$	$= 0.04375 \cdot 15 = 065625$
				$\sum = 4.55$

The joint failure mode method can model dependency between failure modes and account for the combined consequences of joint failures within the system component. As such, this method is the most robust and accurate of the options for system components where joint failures can occur. For example, with many dams there is a possibility that multiple gates can fail during the same flood event, leading increased discharge and flood inundation. While the joint failure mode option is robust, it is also the most computationally intensive of the three options. In RMC-TotalRisk, due to runtime and virtual memory limitations, if the joint failure mode option is selected, the maximum number of failure modes allowable for a single system component is 20. The user should expect noticeably slower runtimes when the number of failure modes exceeds 10.

Multiple System Components

RMC-TotalRisk can perform risk analysis for a single system component, such as a dam or levee, or a complex system with multiple components, where each component can have multiple failure modes. For illustration, a risk analysis in RMC-TotalRisk with two separate dams each with two failure modes is shown in Figure 39 below.

Computing risk for multiple system components requires a multidimensional integral. Consider a system with two components, where the consequences of failure from each component are additive. Following the general risk formula provided in Equation 49, the system risk becomes a two-dimensional integral:

$$\mathbb{E}[C]_{\Omega} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{C_X(x) + C_Y(y)\}\cdot f_{XY}\big(C_X(x), C_Y(y)\big)\cdot dx\cdot dy \qquad \textit{Equation 101}$$

where x is the hazard level for system component X; $C_X(x)$ determines the consequences for the hazard level x; y is the hazard level for system component Y; $C_Y(y)$ determines the consequences for the hazard level y; and $f_{XY}\big(C_X(x), C_Y(y)\big)$ is the joint PDF of the combined system consequences occurring.

Rewriting this as the discrete form of the risk equation shown in Equation 51, the system risk of failure becomes:

$$\mathbb{E}[C_F]_{\Omega} = \sum_i\sum_j P_{XY}\left(x_i, y_j\right)\cdot\left\{P(F_X|x_i)\cdot C_{F_X}(x_i) + P\left(F_Y|y_j\right)\cdot C_{F_Y}\left(y_j\right)\right\} \qquad \textit{Equation 102}$$

where $P_{XY}\big(x_i, y_j\big)$ is the joint probability of the hazard level x_i and y_j; $P(F_X|x_i)$ is the combined conditional probability of failure for system component X given the hazard level x_i; $C_{F_X}(x_i)$ is the consequence of failure for system component X given the hazard level x_i; $P\big(F_Y|y_j\big)$ is the combined conditional probability of failure for system component Y given the hazard level y_j; and $C_{F_Y}\big(y_j\big)$ is the consequence of failure for system component Y given the hazard level y_j.

As shown in Equation **101**Equation 102, the dependency between system components is modeled by the joint probability of the hazard functions $P_{XY}\big(x_i, y_j\big)$. Dependency between system component hazard functions can be set as perfectly independent, positive, or negatively dependent. There is also an option to set the dependency between system components with a user-defined correlation matrix (see **Appendix F**).

The joint failure mode approach in the previous section is used to estimate the combined consequences of failure and non-failure of the system. The user must also select a joint consequence rule: additive, average, maximum, or minimum. In the current version of RMC-TotalRisk, the failure modes within a system component are statistically independent from failure modes within all other system components. For example, the internal erosion failure at Dam A is independent of the internal erosion failure mode at Dam B. Considering this, Equation 102 needs to be expanded to include all combinations of failure as follows:

$$\begin{aligned}\mathbb{E}[C_F]_{\Omega} = &\sum_i\sum_j P_{XY}\left(x_i, y_j\right)\\ &\cdot\left\{P\left(F_X|x_i\cap\overline{F_Y|y_j}\right)\cdot C_{F_X}(x_i) + P\left(\overline{F_X|x_i}\cap F_Y|y_j\right)\cdot C_{F_Y}\left(y_j\right)\right.\\ &\left.+ P\left(F_X|x_i\cap F_Y|y_j\right)\cdot\left[C_{F_X}(x_i) + C_{F_Y}\left(y_j\right)\right]\right\}\end{aligned} \qquad \textit{Equation 103}$$

The formulation in Equation 103 for accounting for joint failures is consistent with what has been presented in [5] and [45]. Computing system risk is computationally demanding. If traditional numerical integration techniques were used, the solution to Equation 103 would require K^D iterations, where K is the number of integration steps (or bins) and D is the number of system components. If there were 100 integration steps and 5 system components, the solution would need 10 billion iterations. To avoid these computational limitations, RMC-TotalRisk uses an adaptive importance sampling method for computing system risk. Please see **Appendix D** for more details.

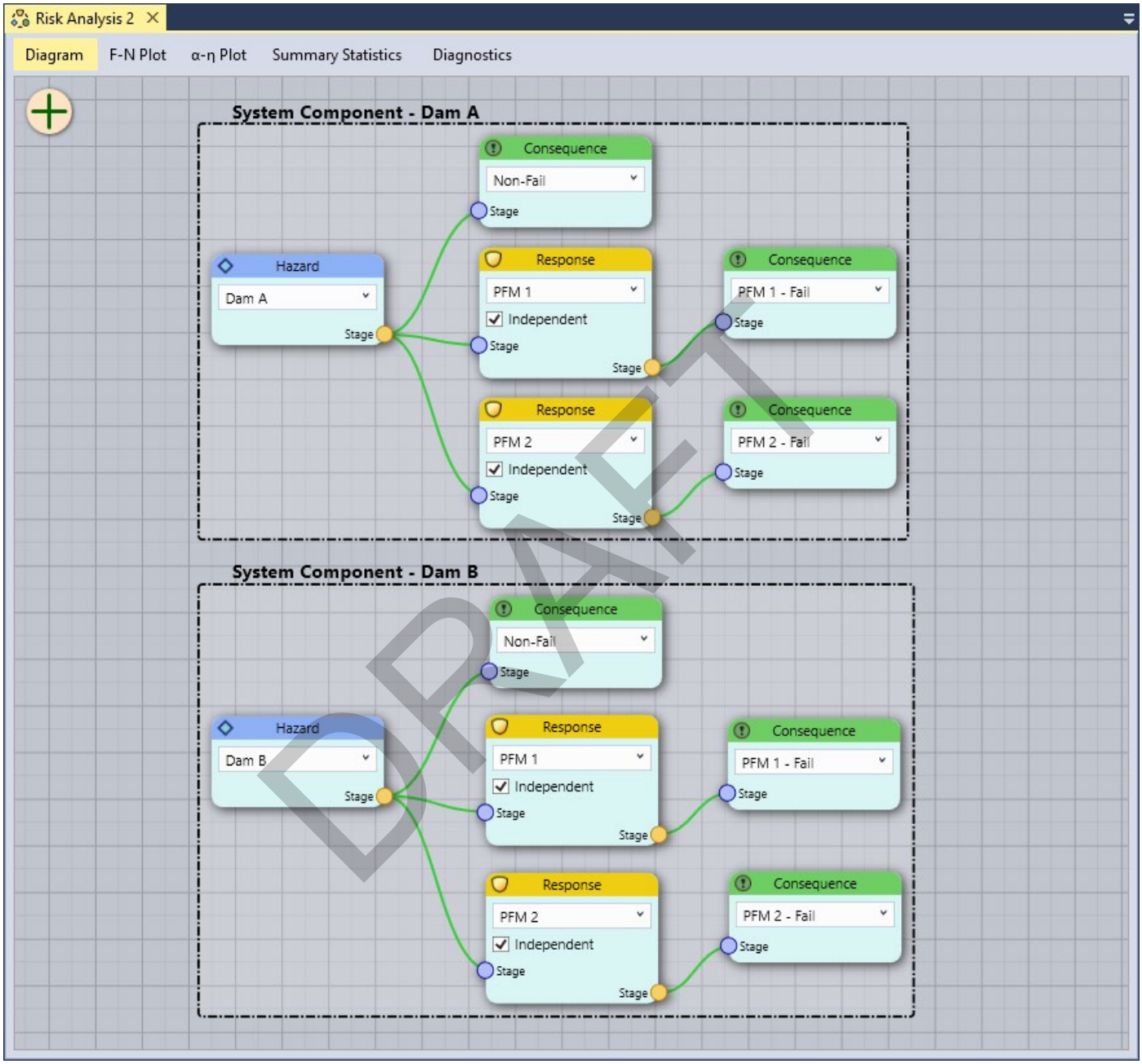


Figure 39 - RMC-TotalRisk risk diagram with two system components.

Simulation Framework

After the inputs have been defined in the risk diagram, the marginal risk for the individual failure modes, risk for the system components, and risk for the overall system can be computed. There are two simulation options: 1) Simulate Mean Risk Only, and 2) Simulate Risk with Full Uncertainty. The following subsections describe these options in greater detail.

Mean Risk Only

There are two primary components of randomness in a quantitative risk analysis: natural variability and knowledge uncertainty. Natural variability (NV) is best described as the effect of randomness and is a function of the system [8]. It is not reducible through either study or further measurement. Knowledge uncertainty (KU) is the lack of knowledge about the parameters or processes that characterize the system being modeled. It can be reduced through further measurement or study. The expected consequences estimated from a quantitative risk analysis are a function of both NV and KU of the model inputs:

$$\mathbb{E}[C] = f(\boldsymbol{NV}, \boldsymbol{KU}) \qquad \textit{Equation 104}$$

For the *mean risk only* option, the effects of NV and KU are quantified by using the expected values of each input function in the overall risk analysis. For example, the expected consequences of failure considering both NV and KU in each function are estimated as follows:

$$\mathbb{E}[C_F] = \sum_i \mathbb{E}[P(x_i)] \cdot \mathbb{E}[P(F|x_i)] \cdot \mathbb{E}[C_F(x_i)] \qquad \textit{Equation 105}$$

where $\mathbb{E}[P(x_i)]$ is the expected probability of the hazard level x_i; $\mathbb{E}[P(F|x_i)]$ is the expected conditional probability of failure given the hazard level x_i.; and $\mathbb{E}[C_F(x_i)]$ is the expected consequence of failure given the hazard level x_i.

Equation 105 relies on the *law of total expectation*, which states that when random variables are statistically independent, the expectation of their product is the product of their expectations:

$$\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y] \qquad \textit{Equation 106}$$

Using this simplifying assumption, simulating the *mean risk only* can be performed very efficiently with a single loop, as shown on the left side of Figure 40 below. In this figure, the white rectangles indicate where KU is accounted for, and the rounded light blue rectangles indicate where NV is accounted for.

For the *mean risk only* option to be accurate, the mean hazard function and mean response function need to provide the expected probability given the hazard level, $\mathbb{E}[p|x]$. This concept was first presented by Beard [46] as the *expected probability of exceedance*. The expected probability given the hazard level $\mathbb{E}[p|x]$ is not equal to the expected hazard given the probability $\mathbb{E}[x|p]$, due to asymmetries in NV and KU of the parent probability distribution.

The *mean risk only* option will generally produce accurate estimates of the mean risk. For example, if the system response function is an event tree, the *mean risk only* option provides an exact estimate of $\mathbb{E}[P(F|x_i)]$. Whereas the full uncertainty option will have some minor Monte Carlo sampling errors when multiplying small probabilities in event trees. However, the the *mean risk only* option cannot capture all the combinations of low probability and high consequence events, which can lead to an underestimation of the tail risk. For scenarios where there is a lot of uncertainty in the extremes, it is recommended to simulate risk with full uncertainty.

Flood risk management (FRM) decisions are commonly made based on the mean value of risk. The federal guidance for the economic evaluation of FRM studies is to use expected annual damages (or life loss) as the primary decision criterion [7] [11]. USACE regulation [11] requires dam safety alternatives to be evaluated using a cost-effectiveness measure called the cost-to-save-a-statistical-life (CSSL), which is estimated using expected annual life loss (lives/year) and expected annual economic damages ($/year). Considering these policy requirements, the *mean risk only* option provides an efficient and accurate simulation method for comparing risk reduction alternatives.

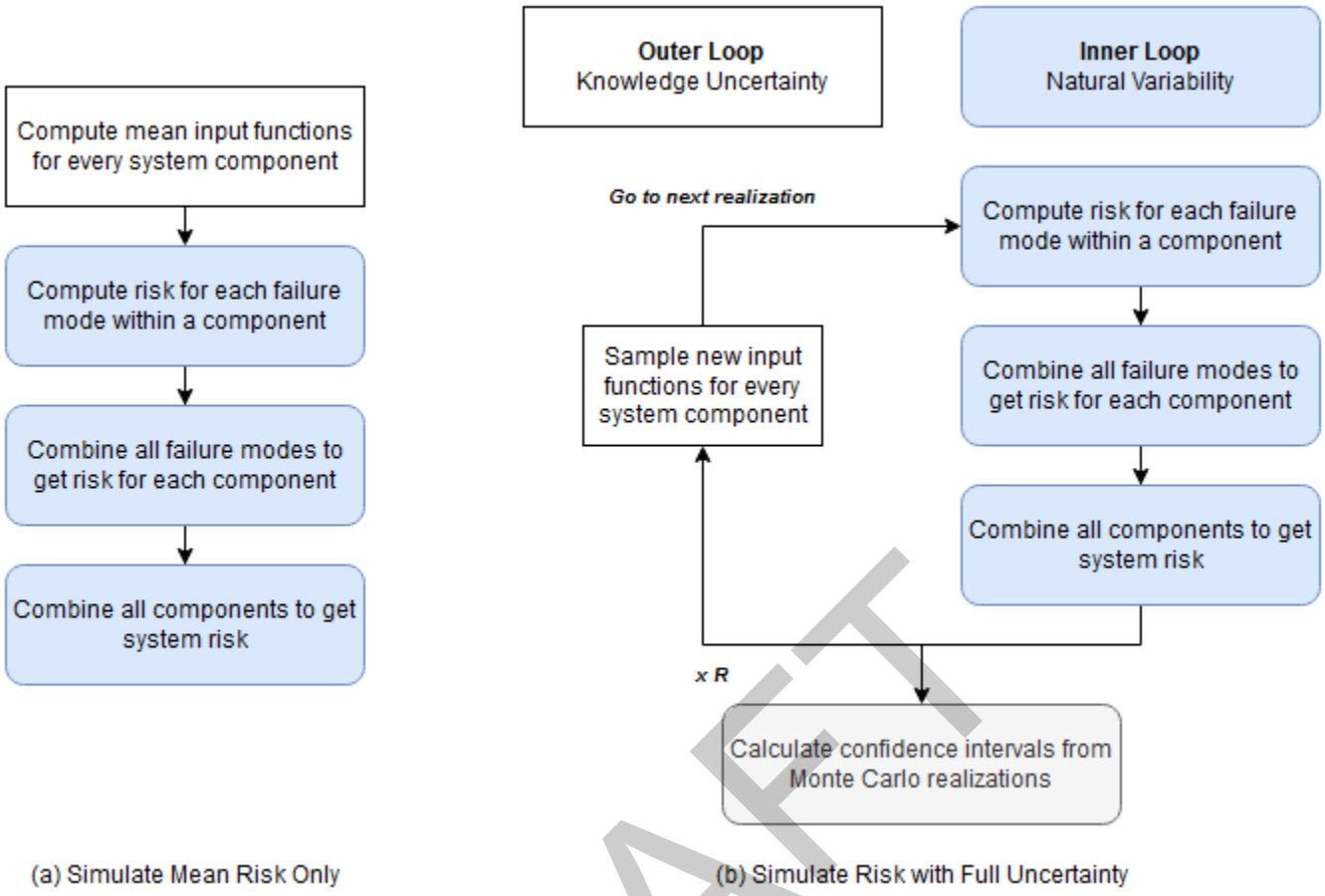


Figure 40 - Flowchart of the RMC-TotalRisk simulation options: (a) Simulate mean risk only, and (b) Simulate risk with full uncertainty.

Risk with Full Uncertainty

The second option in RMC-TotalRisk is to simulate risk with full uncertainty. This option requires much more computational effort, but it allows for the portrayal of KU and the development of confidence intervals on the risk results. If the risk analysts and decision makers are interested in exploring and evaluating ways to reduce KU, it is recommended to simulate risk with full uncertainty.

Simulating risk with full uncertainty requires two loops, as shown on the right side of Figure 40. This Monte Carlo simulation approach is sometimes referred to as *two-dimensional simulation* [47] or *two looped nested Monte Carlo* [12]. For each Monte Carlo realization, new input functions are randomly sampled for every system component, following the procedures described in the respective input function chapters. For example, Algorithm 2 describes the parametric hazard function uncertainty analysis. Given the new set of input functions, risk is computed for each failure mode for each component in the system. After the simulation is complete, confidence intervals can be calculated using the results from all the Monte Carlo realizations.

The full uncertainty simulation is more formally described in Algorithm 8 below. The user must enter the number of Monte Carlo realizations. The default is 1,000 to ensure reasonably accurate confidence intervals and shorter runtimes. Due to runtime and file size limitations, the maximum number of Monte Carlo realizations allowed is 10,000. As such, there is a potential for minor sampling errors in the mean risk results and percentiles. Please see **Appendix C** for details on Monte Carlo convergence.

Algorithm 8 – Simulate Risk with Full Uncertainty

$R \leftarrow$ number of Monte Carlo realizations
$D \leftarrow$ number of system components
for $i \leftarrow 1$ to R **do**
  for $j \leftarrow 1$ to D **do**
    $n_F \leftarrow$ number of input functions in component j
    for $k \leftarrow 1$ to n_F **do**
      Sample new input function $F^*_{i,j,k}$ given random percentile $r_k \sim \mathrm{U}(0,1)$.
    end for
    $n_{fm} \leftarrow$ number of failure modes in component j
    for $k \leftarrow 1$ to n_{fm} **do**
      Compute risk of failure $\mathbb{E}[C_F]^*_{i,k}$ for failure mode k using new input functions $\boldsymbol{F}^*_{i,j}$
      $\mathbb{E}[C_F]^*_{i,j} \leftarrow \mathbb{E}[C_F]^*_{i,j} + \mathbb{E}[C_F]^*_{i,k}$ ⊳Keep track of component risk of failure
    end for
    Compute risk of non-failure $\mathbb{E}[C_{NF}]^*_{i,j}$ for component j
    Compute total risk $\mathbb{E}[C_T]^*_{i,j} = \mathbb{E}[C_F]^*_{i,j} + \mathbb{E}[C_{NF}]^*_{i,j}$ for component j
    $\mathbb{E}[C_T]^*_{\Omega_i} \leftarrow \mathbb{E}[C_T]^*_{\Omega_i} + \mathbb{E}[C_T]^*_{i,j}$ ⊳Keep track of system risk
    ⊳Note: This pseudo-code is for demonstration. In RMC-TotalRisk, all risk types are computed and recorded for each failure mode, each component, and the overall system.
  end for
  Record all risk results for each risk type, component, and failure mode into a data frame:
  $\boldsymbol{\theta}^*_i \leftarrow \left\{\mathbb{E}[C_T]^*_{\Omega_i}, \mathbb{E}[C_T]^*_{i,1}, \cdots, \mathbb{E}[C_T]^*_{i,D}\right\}$
end for
Estimate confidence intervals and mean risk results from realizations $\{\boldsymbol{\theta}^*_1, \cdots, \boldsymbol{\theta}^*_R\}$.

Risk Results

In RMC-TotalRisk, there are three main ways to view the risk results: 1) the F-N plot; 2) the α-η plot; and 3) the summary statistics table. Each of these are described in more detail in the following subsections.

F-N Plot

The F-N plot has consequences (N) on the x-axis and the exceedance probability, or frequency of occurrence, (F) on the y-axis [48]. This type of plot is also commonly referred to as an *F-N curve*, *survival function*, or *Farmer diagram.* A log-log scale is typically used because the range of probabilities and consequences can span multiple orders of magnitude.

The F-N curve is constructed by first sorting the risk results by decreasing values of consequences (N). Then, the probability of the consequences occurring (e.g., $P(x_i) \cdot P(F|x_i)$) are summed to get the exceedance probability (or cumulative frequency). The overall risk integral is often evaluated at hundreds or even thousands of points during the risk compute. To reduce memory requirements in the software, the F-N curves are thinned down based on a user-specified F-N output length, which by default is 200 points.

An example F-N plot is provided in Figure 41 below. This example shows results from a risk analysis for a dam with both seismic and hydrologic failure modes. The incremental risk from seismic related failure modes is shown in red, the incremental risk from hydrologic failure modes in blue, and the combined incremental risk in black. As shown on the left side of the figure, results can be filtered by risk type and by system components and individual failure modes.

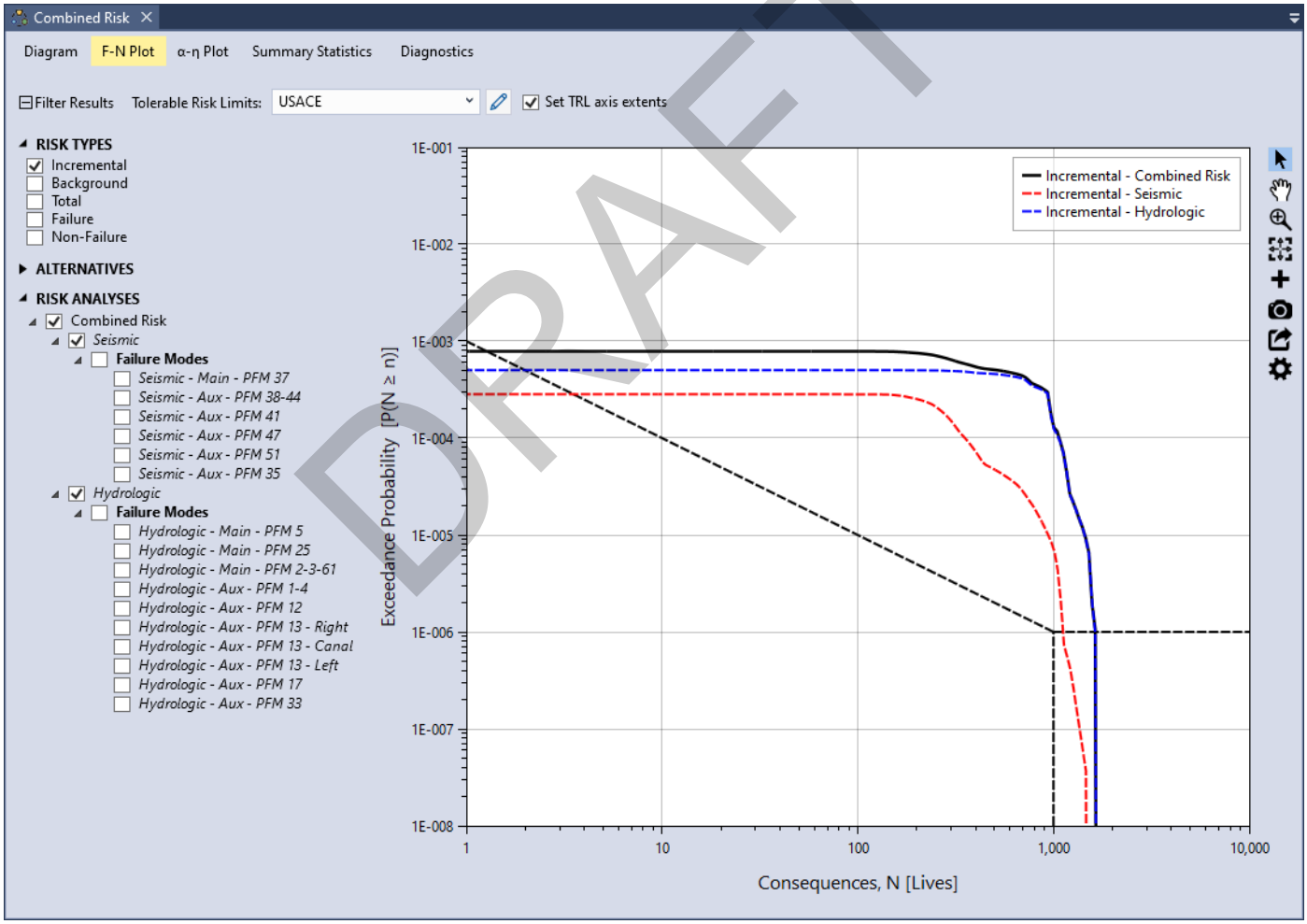


Figure 41 - Example of F-N plot results.

By default, the USACE tolerable risk limit [11] is also plotted. The tolerable risk limit (or guideline) can be customized or removed from the plot. Figure 42 shows an F-N plot for another example risk analysis after simulating risk with full uncertainty. The 90% confidence interval is shown as a shaded bound. The mean and median curves are also provided.

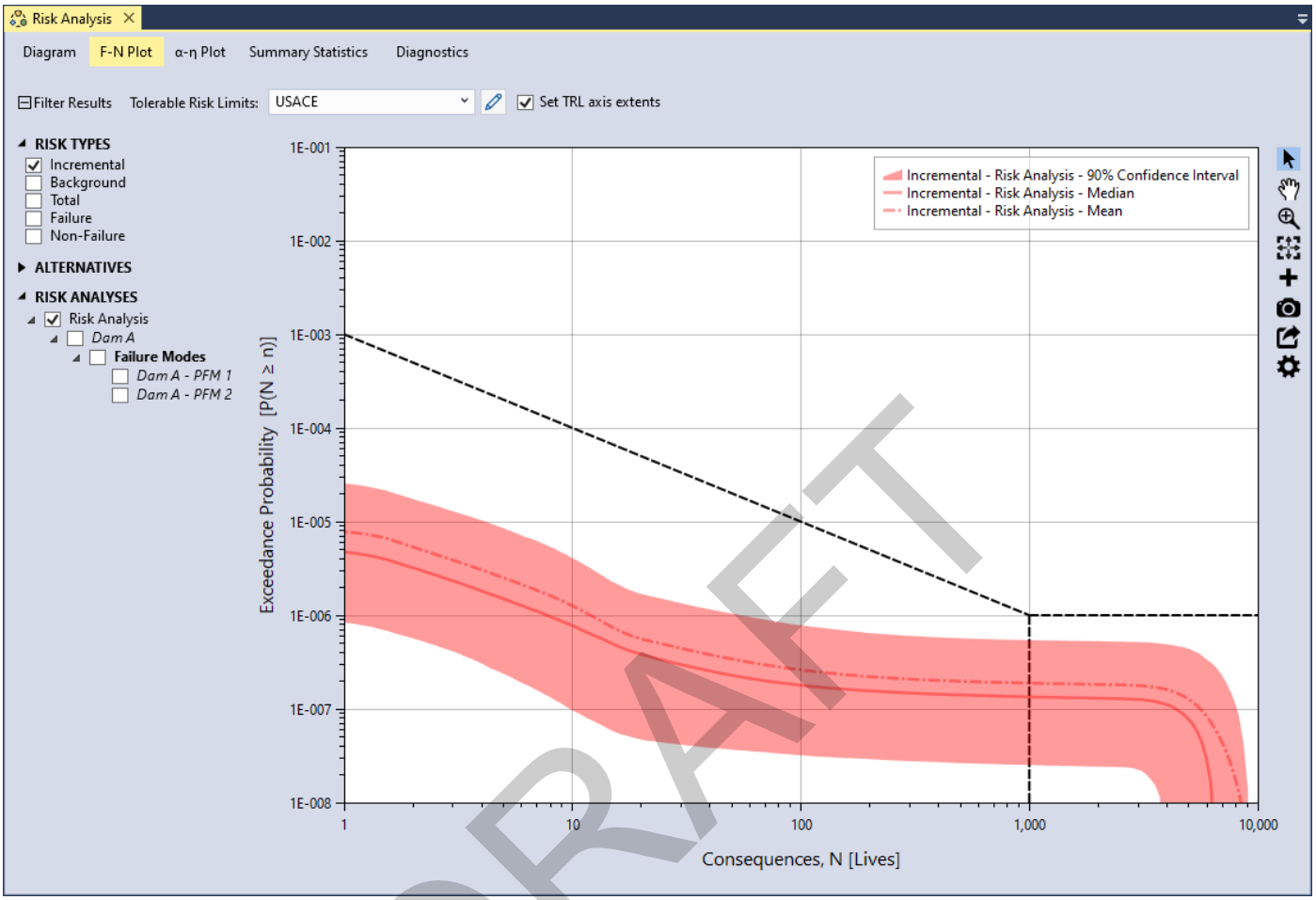


Figure 42 - Example of F-N plot results with confidence intervals.

α-η Plot

Risk results can be viewed with the α-η plot (pronounced as “alpha-N”), which is a commonly used in the USACE dam and levee safety programs for plotting incremental risk. In this plot, the conditional mean consequences (η) are plotted on the x-axis and the exceedance probability (α) on the y-axis. Like the F-N plot, a log-log scale is typically used because the range of probabilities and consequences can span multiple orders of magnitude.

For the risk of failure and incremental risk, the conditional expectation is implicitly conditional on the critical hazard level x_c beyond which there is a nonzero probability of failure $P(F|x_c) > 0$. For example, the conditional expectation (η) for incremental risk is computed as follows:

$$\eta_\Delta = \mathbb{E}[C_\Delta | x \geq x_c] = \frac{\sum_{i=x_c}^{n} P(x_i) \cdot P(F|x_i) \cdot C_\Delta(x_i)}{\sum_{i=x_c}^{n} P(x_i) \cdot P(F|x_i)} = \frac{\mathbb{E}[C_\Delta]}{\alpha_\Delta} \qquad \textit{Equation 107}$$

where $C_\Delta(x_i) = C_F(x_i) - C_{NF}(x_i)$; x_c is the threshold hazard level where there is a nonzero probability of failure; and α_Δ is the exceedance probability for which incremental consequences would occur.

The product of the exceedance probability and the conditional mean is equal to the unconditional mean:

$$\mathbb{E}[C_\Delta] = \alpha_\Delta \cdot \eta_\Delta \qquad \textit{Equation 108}$$

In the USACE dam and levee safety programs, Equation 108 has traditionally been written as $f \cdot \bar{N}$:

$$\mathbb{E}[C_\Delta] = f_\Delta \cdot \bar{N}_\Delta \qquad \textit{Equation 109}$$

However, in probability and statistics, f is universally used as the symbol for the probability density function, not an exceedance probability. Therefore, to avoid any unnecessary confusion and to be consistent with other disciplines, RMC-TotalRisk uses α as the symbol for exceedance probability.

The α-η plot is only available for the risk of failure and incremental risk. Both risk types have the same exceedance probability, $\alpha_F = \alpha_\Delta$. Total risk and background risk are both unconditional expectations where $\alpha = 1$. The risk of non-failure is a conditional expectation, where $\alpha_{NF} = 1 - \alpha_F$. However, α_{NF} is typically very close to 1, so there is no reason to plot it.

An example α-η plot is shown in Figure 43 below. As with the previous F-N plot example, Figure 43 shows the incremental risk from seismic related failure modes in red, the incremental risk from hydrologic failure modes in blue, and the combined incremental risk in black. The combined incremental risk has an exceedance probability $\alpha_\Delta = 7.8626e^{-4}$ and conditional mean incremental consequences $\eta_\Delta = 700.2036$. Stated another way, the annual probability of failure is $7.8626e^{-4}$, and if the dam were to fail, the expected incremental life loss would be 700.2036.

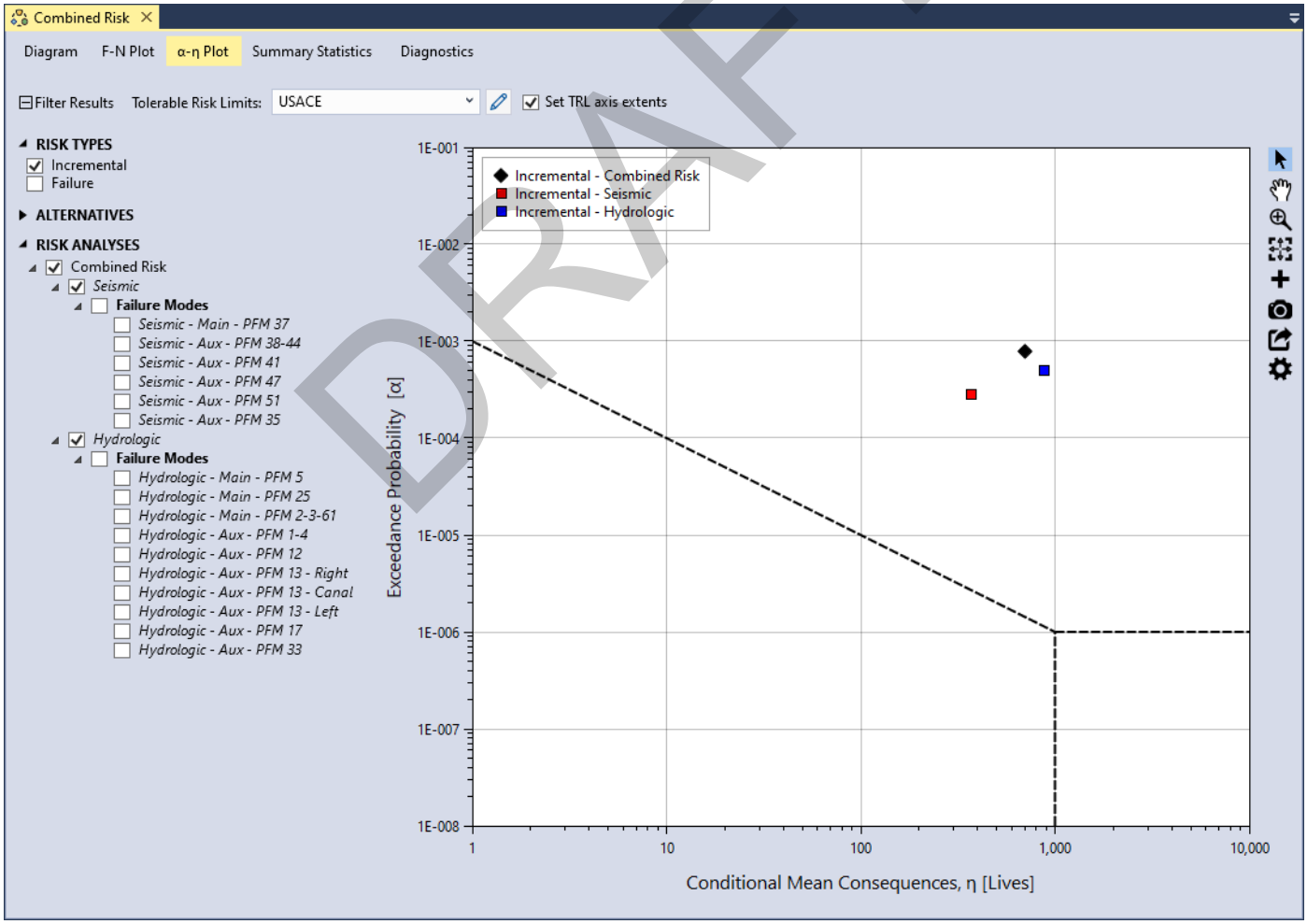


Figure 43 – Example of α-η plot results.

The diagonal of the α-η plot is equal to the product of α and η, which is the unconditional mean consequences, $\mathbb{E}[N] = \alpha \cdot \eta$. The α-η points will typically plot near the inflection point of the F-N curve, with α equaling the maximum exceedance probability from the F-N curve.

Figure 44 shows an α-η plot after simulating risk with full uncertainty. The uncertainty is portrayed as a scatter cloud. To improve the visibility of this plot, the uncertainty scatter is thinned down to ensure a maximum of 1,000 points per system risk component.

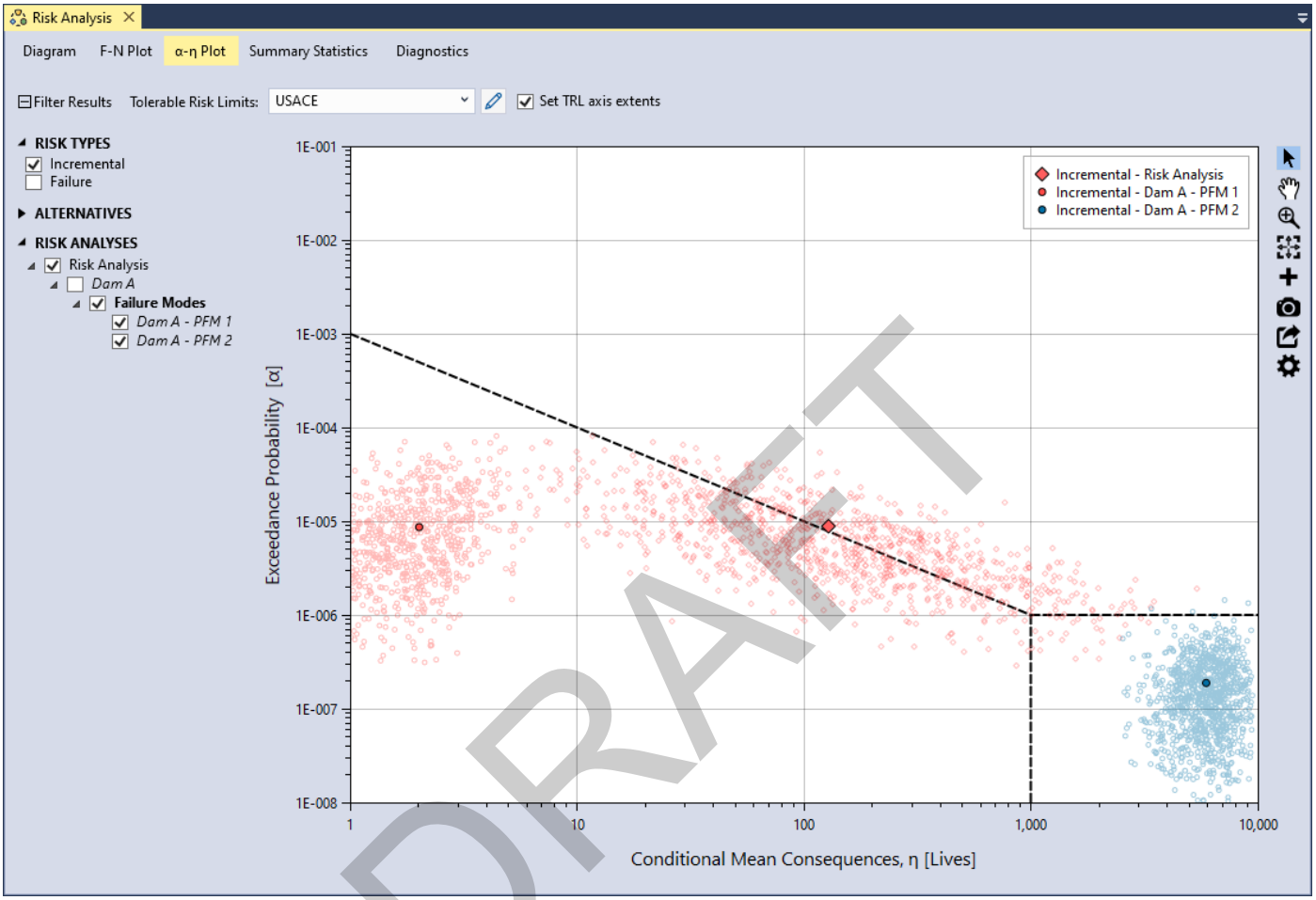
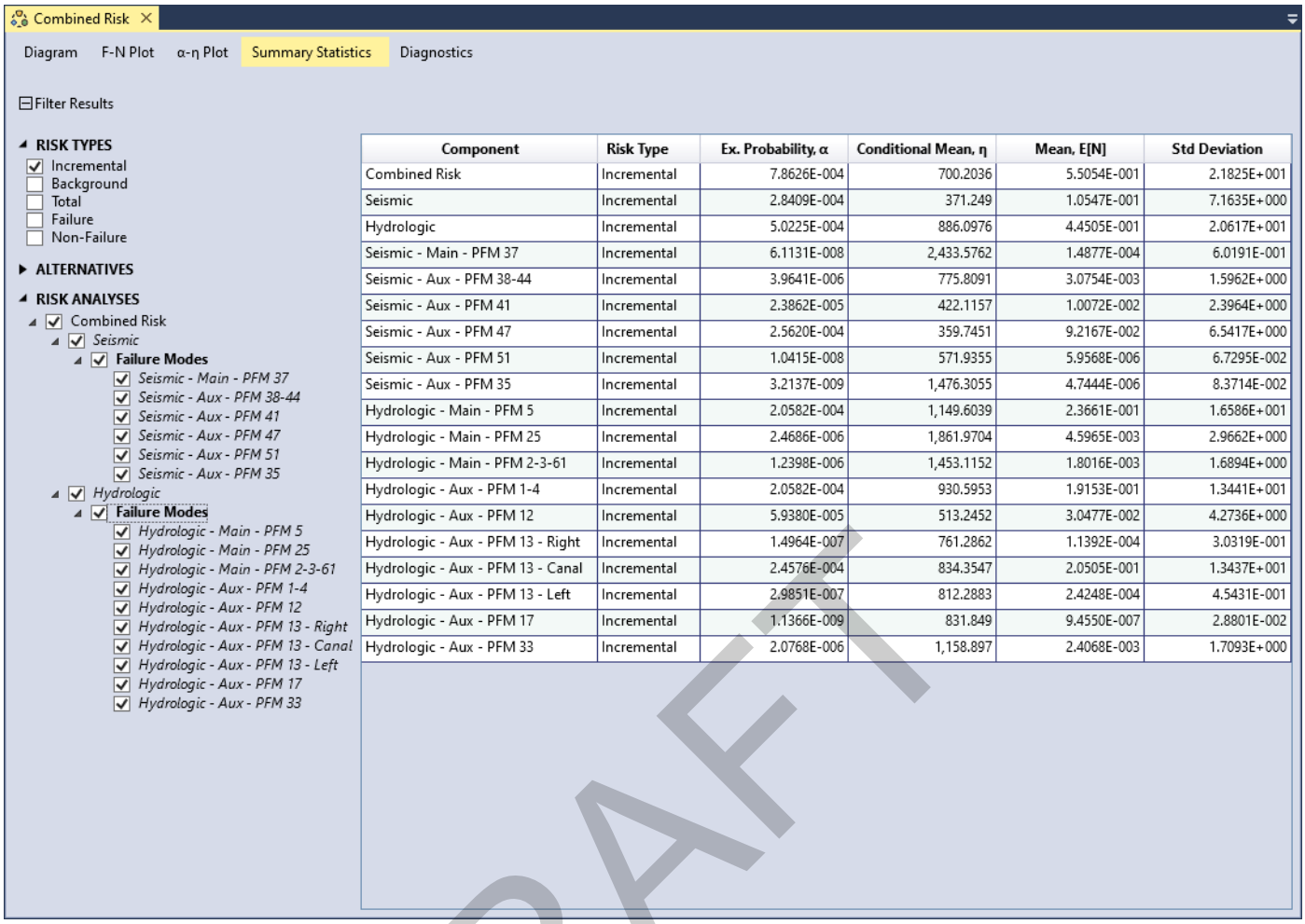


Figure 44 – Example of α-η plot results with uncertainty scatter.

Summary Statistics

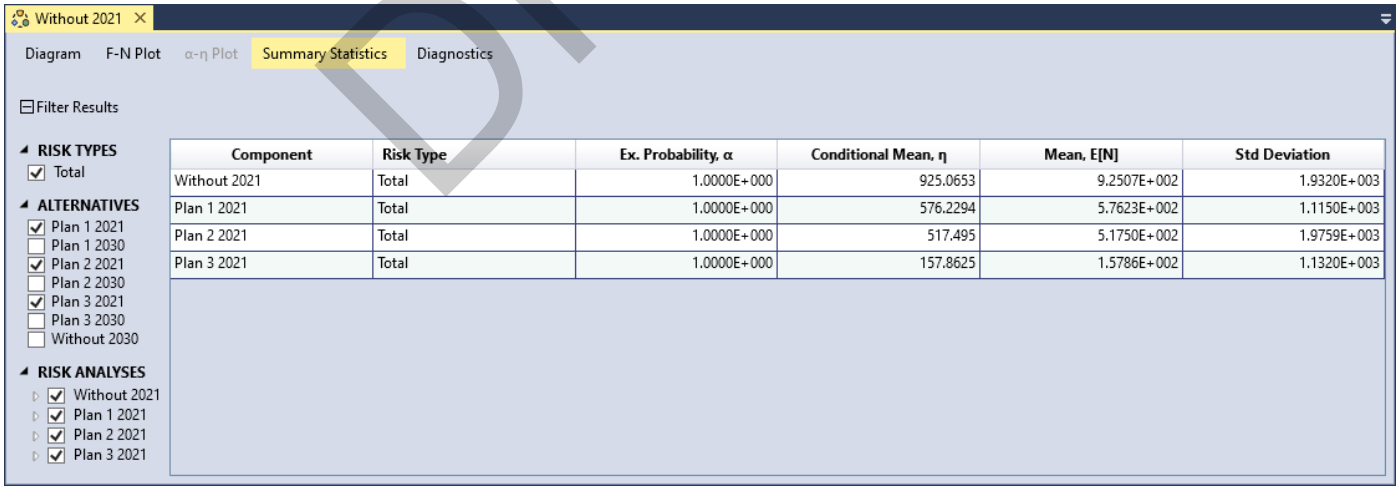
Summary statistics are provided for each risk type being evaluated as shown in Figure 45 below. Statistics are provided for each failure mode for each system component, as well as for the full system. In the example shown in Figure 45, the probability of failure for the dam and each failure mode is provided in the column labelled "**Ex. Probability, α**". The expected incremental consequences given failure are provided in the "**Conditional Mean, η**" column. The expected annual incremental consequences are provided in the "**Mean, E[N]**" column. The product of α and η is equal to the mean, $\mathbb{E}[N] = \alpha \cdot \eta$. Finally, the standard deviation of the incremental consequences is provided in the "**Std. Deviation, σ**" column. More details on these summary statistics are provided in **Appendix B**.

The summary statistics for a risk analysis can be compared with other alternative analyses within the same study. Figure 46 below shows an example where the risk at a river reach without a new levee is compared to three alternative plans that include construction of a new levee. In this example, the total risk is displayed, therefore the exceedance probability $\alpha = 1$, the conditional mean equals the unconditional mean, $\mathbb{E}[N] = \eta$.



Component	Risk Type	Ex. Probability, α	Conditional Mean, η	Mean, E[N]	Std Deviation
Combined Risk	Incremental	7.8626E-004	700.2036	5.5054E-001	2.1825E+001
Seismic	Incremental	2.8409E-004	371.249	1.0547E-001	7.1635E+000
Hydrologic	Incremental	5.0225E-004	886.0976	4.4505E-001	2.0617E+001
Seismic - Main - PFM 37	Incremental	6.1131E-008	2,433.5762	1.4877E-004	6.0191E-001
Seismic - Aux - PFM 38-44	Incremental	3.9641E-006	775.8091	3.0754E-003	1.5962E+000
Seismic - Aux - PFM 41	Incremental	2.3862E-005	422.1157	1.0072E-002	2.3964E+000
Seismic - Aux - PFM 47	Incremental	2.5620E-004	359.7451	9.2167E-002	6.5417E+000
Seismic - Aux - PFM 51	Incremental	1.0415E-008	571.9355	5.9568E-006	6.7295E-002
Seismic - Aux - PFM 35	Incremental	3.2137E-009	1,476.3055	4.7444E-006	8.3714E-002
Hydrologic - Main - PFM 5	Incremental	2.0582E-004	1,149.6039	2.3661E-001	1.6586E+001
Hydrologic - Main - PFM 25	Incremental	2.4686E-006	1,861.9704	4.5965E-003	2.9662E+000
Hydrologic - Main - PFM 2-3-61	Incremental	1.2398E-006	1,453.1152	1.8016E-003	1.6894E+000
Hydrologic - Aux - PFM 1-4	Incremental	2.0582E-004	930.5953	1.9153E-001	1.3441E+001
Hydrologic - Aux - PFM 12	Incremental	5.9380E-005	513.2452	3.0477E-002	4.2736E+000
Hydrologic - Aux - PFM 13 - Right	Incremental	1.4964E-007	761.2862	1.1392E-004	3.0319E-001
Hydrologic - Aux - PFM 13 - Canal	Incremental	2.4576E-004	834.3547	2.0505E-001	1.3437E+001
Hydrologic - Aux - PFM 13 - Left	Incremental	2.9851E-007	812.2883	2.4248E-004	4.5431E-001
Hydrologic - Aux - PFM 17	Incremental	1.1366E-009	831.849	9.4550E-007	2.8801E-002
Hydrologic - Aux - PFM 33	Incremental	2.0768E-006	1,158.897	2.4068E-003	1.7093E+000

Figure 45 - Example of summary statistics results.



Component	Risk Type	Ex. Probability, α	Conditional Mean, η	Mean, E[N]	Std Deviation
Without 2021	Total	1.0000E+000	925.0653	9.2507E+002	1.9320E+003
Plan 1 2021	Total	1.0000E+000	576.2294	5.7623E+002	1.1150E+003
Plan 2 2021	Total	1.0000E+000	517.495	5.1750E+002	1.9759E+003
Plan 3 2021	Total	1.0000E+000	157.8625	1.5786E+002	1.1320E+003

Figure 46 - Example of comparing alternative summary statistics results.

Diagnostics

RMC-TotalRisk provides several diagnostics for exploring the Monte Carlo simulation results for a risk analysis. If no uncertainty has been defined in the risk analysis inputs, the diagnostic tools provide limited value. However, if uncertainty has been defined, the diagnostic features include the following:

- **Integration**: The integration diagnostics include an estimate of the standard error of the computed total risk for the system, and the number of integrand function evaluations performed during the risk simulation. The diagnostics are graphically displayed as a kernel density or cumulative distribution plot. Summary statistics, including the mean, standard deviation, and key percentiles are provided in a table. Details on summary statistics and kernel density estimation are provided in **Appendix B**. Details on numerical integration are provided in **Appendix D**.

- **Risk Measures**: Six risk measures are computed for all five risk types. A kernel density plot and cumulative distribution plot are provided to understand the shape and distribution of various risk measures. Summary statistics for each risk measure are provided in a table. More details on additional risk measures are provided in the next section.

- **Risk Profile**: A risk profile plots the exceedance probabilities or conditional mean consequences against increasing hazard levels. The risk profile results can be filtered by system component and risk type. This plot is useful for identifying critical hazard levels where the probability of failure or risk sharply increases.

- **Assurance**: An assurance plot and summary statistics are provided based on a user-defined profile hazard type and hazard threshold. This diagnostic is intended to support the National Flood Insurance Program for levees. More details are provided in **Appendix I**.

- **Tornado Plot**: A tornado plot is provided for visually assessing how sensitive the risk results are to the input functions at each hazard level. The tornado plot is constructed based on the selected system component, risk type, and hazard level. The inputs are ranked from most sensitive at the top to least sensitive at the bottom. An example of a tornado plot is shown below in Figure 47. Details on the sensitivity analysis is provided in **Appendix G**.

- **X-Y Plot**: The X-Y plot is provided for visually assessing the correlation between the different system risk components. Results can be filtered by the risk type, risk measure, and system risk components. For example, the overall risk of failure at the dam could be the Y parameter, and the risk of failure from an individual failure mode could be the X parameter.

- **Tabular**: Tabular results are provided based on the selected risk type and risk measure. The table has a column for each system component and a row for each Monte Carlo realization. The data in this table can be exported, copied, or analysed using the table column tools. Summary statistics for each column in the table are available by right-clicking the column header. Details on these summary statistics are provided in **Appendix B**.

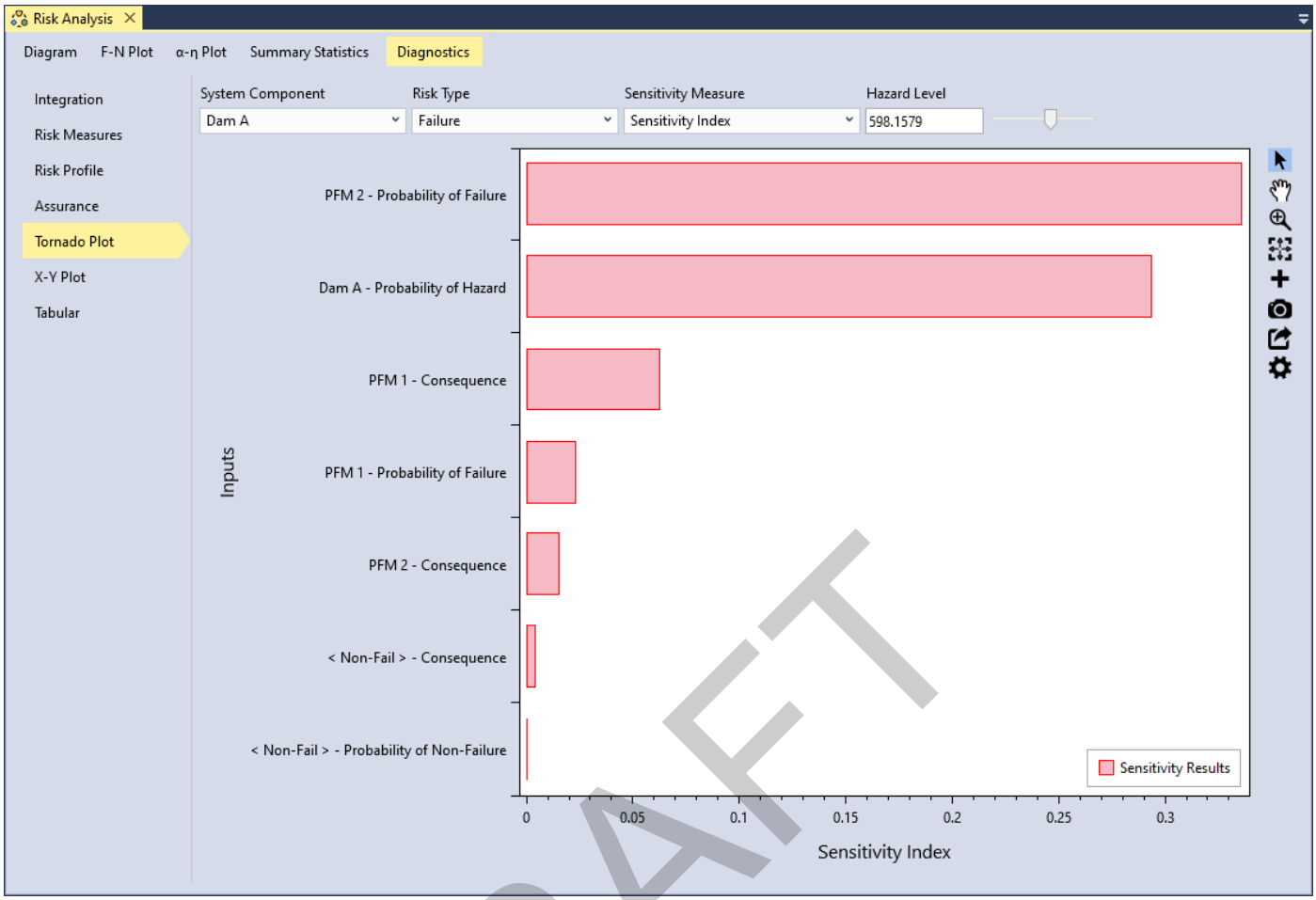


Figure 47 - Example of a tornado plot for risk diagnostics.

Risk Measures

RMC-TotalRisk provides additional risk measures that are useful for risk-based design of engineering structures. For example, a structure might be designed to a specified exceedance probability, e.g., 0.01 or a 100-year level of protection. The following subsections describe the risk measures and how they can be used to inform designs.

Consequence Threshold Probability

The user can enter a consequence threshold n. The probability of consequences exceeding the threshold will be recorded in the risk simulation. The default consequence threshold is 0. The exceedance probability is interpolated from the F-N curve, which is a survival function, for each risk type:

$$S(n) = 1 - F(n) = P(N \geq n) \qquad \textit{Equation 110}$$

After simulating risk with full uncertainty, a kernel density plot and summary statistics are provided as shown in Figure 48 below. The summary statistics are provided in a table on above the plot. The mean results are shown as a vertical dashed line on the plot. This risk measure provides the expected probability of exceedance of a specified consequence level.

When designing a new levee, an objective could be that a consequence threshold of 0 should only be exceeded 1:100 years, i.e., $P(N \geq 0) \leq 0.01$. From Figure 48, the mean threshold probability is $5.3103e^{-3}$ and the 95th percentile is $9.1657e^{-3}$, which demonstrates that the objective is satisfied, with at least 90% confidence.

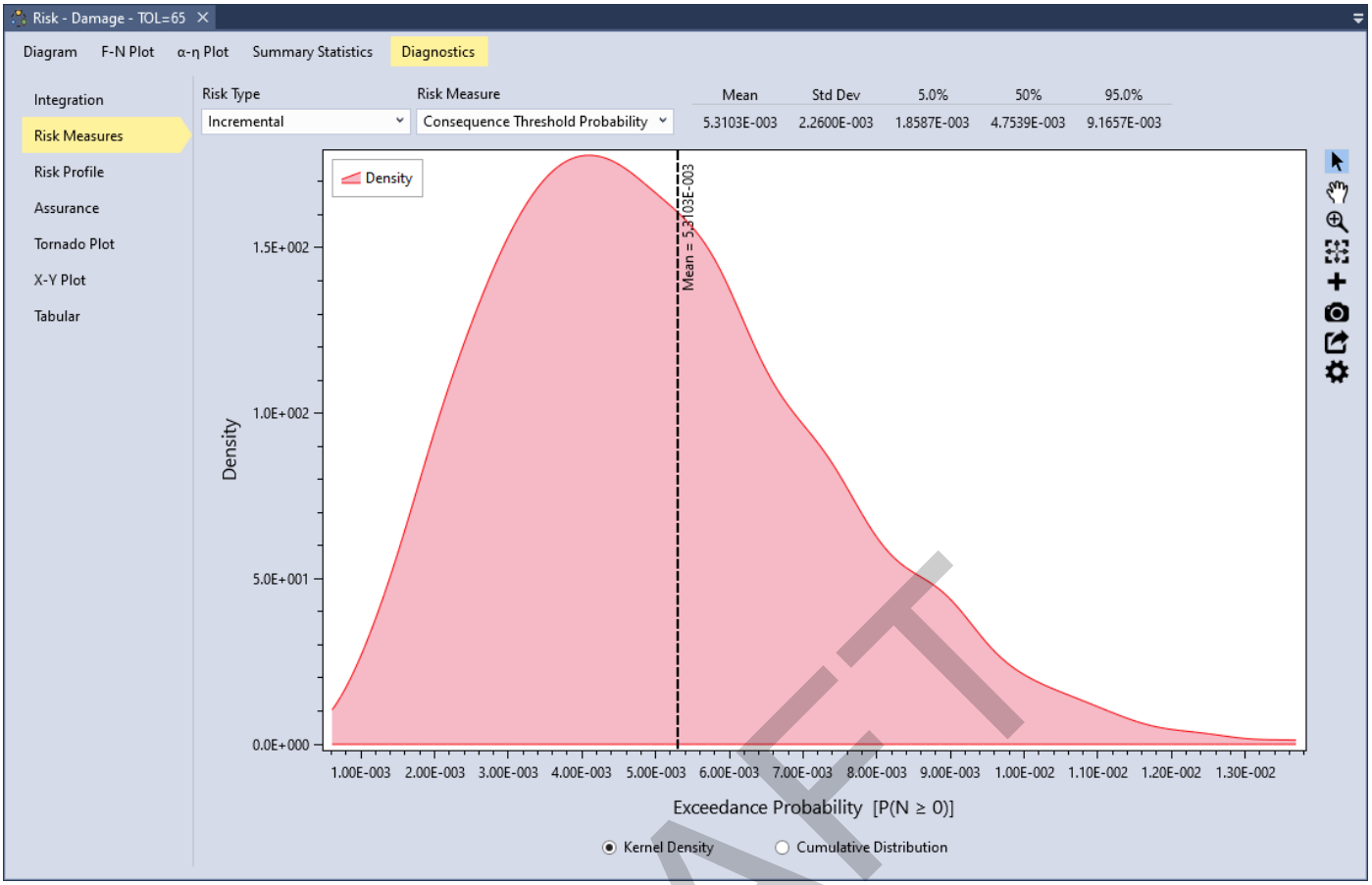


Figure 48 - Example diagnostic plot for risk measures.

Value-at-Risk

The user can enter an exceedance probability α, such as 0.01. This exceedance probability is used for computing the Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR). The default exceedance probability is 0.01. The VaR provides the minimum consequences for the user-specified exceedance probability α. The value is interpolated from the F-N curve for each risk type:

$$S^{-1}(\alpha) = n \qquad \textit{Equation 111}$$

The objective for a new levee could be that the minimum consequences allowable for a 100-year flood must be 0. This risk measure can be used to determine if that design objective has been satisfied with a desired level of confidence.

Conditional Value-at-Risk

The Conditional Value-at-Risk (CVaR) provides the mean (or expected) consequences for the user-specified exceedance probability α. The value is computed by integrating under the F-N curve for each risk type in the prescribed probability range:

$$\mathbb{E}[N|N \geq \beta] = \frac{1}{\alpha}\int_{\beta}^{\infty} x \cdot dS(x) \qquad \textit{Equation 112}$$

where $\beta = S^{-1}(\alpha) = n$. CVaR is estimated using numerical integration with adaptive Simpson's rule (see **Appendix D** for more details).

The objective for a new levee could be that the mean consequences allowable for a 100-year flood must be less than 100. This risk measure can be used to determine if that design objective has been satisfied with a desired level of confidence.

Figure 49 and Figure 50 below illustrate the differences between the expected value (mean), VaR, and CVaR. The mean $\mathbb{E}[\mathrm{N}]$ is the center of mass of the risk distribution. The VaR is evaluated at the exceedance probability α, and represents the minimum consequences at that point. The CVaR provides the average consequences for all events greater than VaR, which is illustrated as the light blue shaded region in both figures.

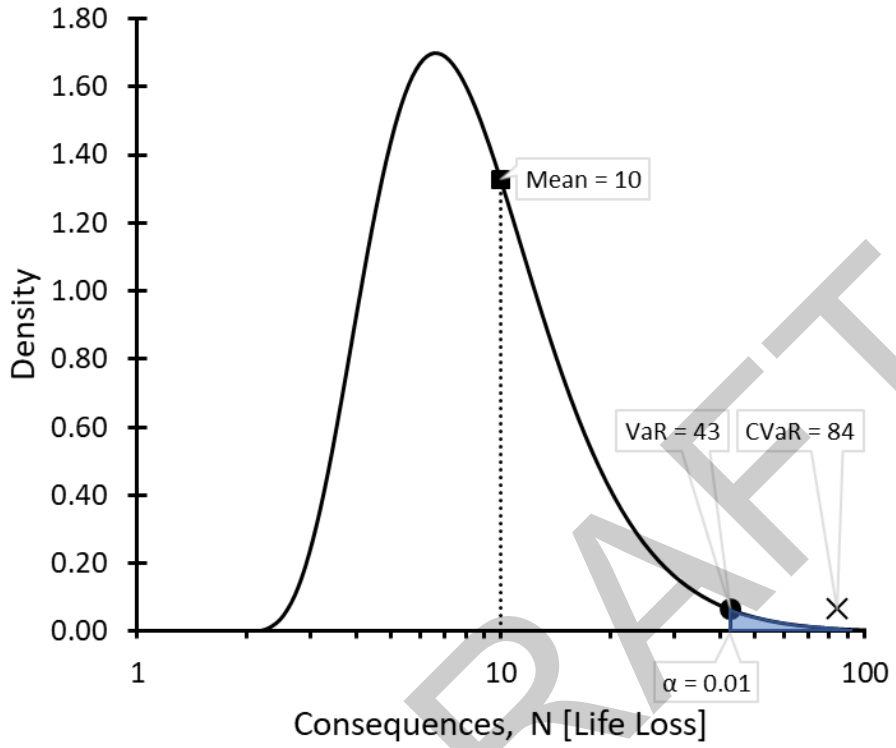


Figure 49 – Density plot illustrating the differences between the expected value (mean), value-at-risk (VaR), and conditional value-at-risk (CVaR).

CVaR measures the expected consequences of an extreme event that exceeds the user-specified exceedance probability α. CVaR was developed in the financial industry to minimize the risk of unacceptable monetary loss in an investment portfolio [49]. Alternative names for CVaR found in literature are *Average Value-at-Risk*, *Expected Shortfall*, and *Expected Tail Loss*. CVaR belongs to a family of coherent risk measures that has stable integral characteristics [50], and is more conservative than the expected value and VaR. Smith [51] demonstrated that CVaR provides a robust risk-based optimization framework for flood risk management studies and provided a case study demonstrating how to use CVaR to optimize the design of a new levee.

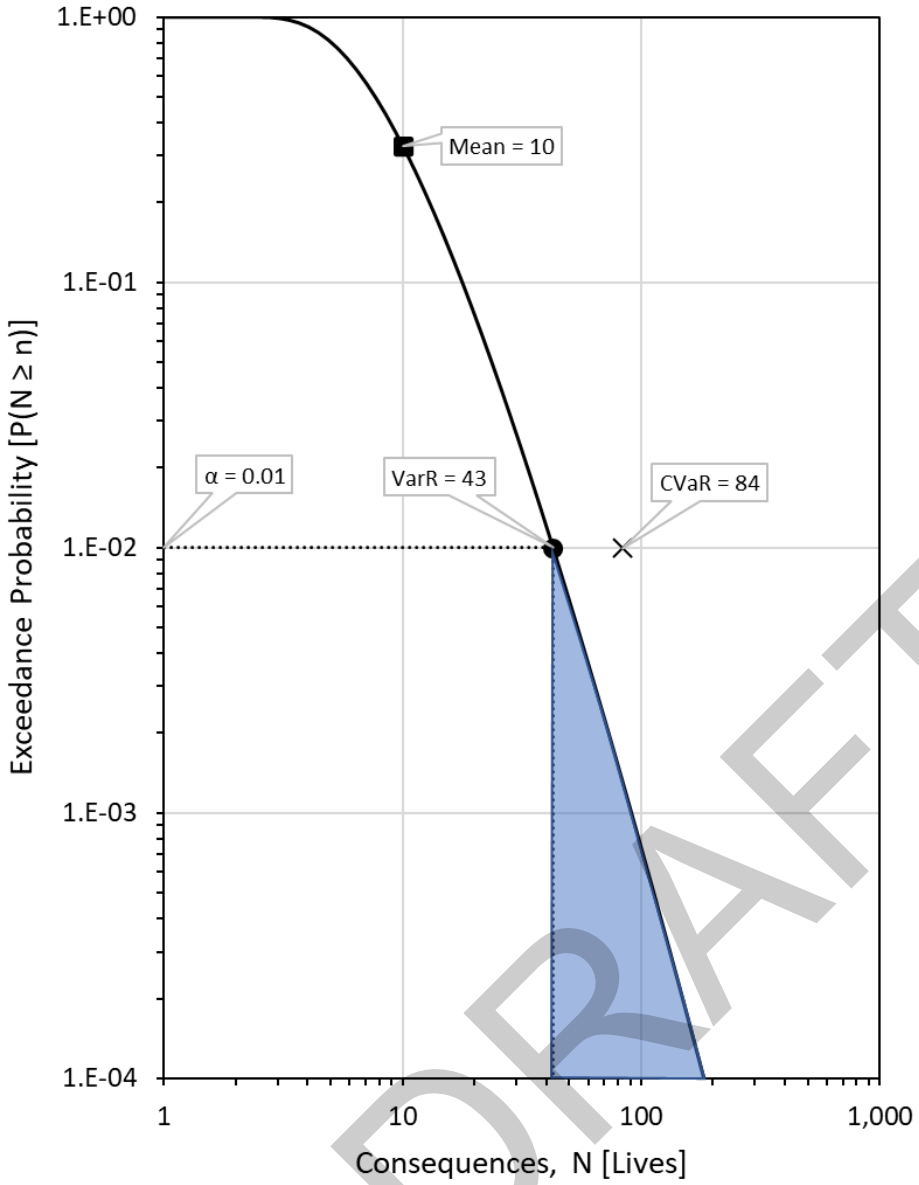


Figure 50 - F-N plot illustrating the differences between the expected value (mean), value-at-risk (VaR), and conditional value-at-risk (CVaR).

References

[1] C. H. Smith, W. L. Fields and N. J. Snorteland, "A New Suite of Risk Analysis Software for Dam and Levee Safety," *The Journal of Dam Safety,* pp. 36-46, 2021.

[2] U.S. Army Corps of Engineers, "ER 1105-2-101 Risk Assessment for Flood Risk Management Studies," U.S. Army Corps of Engineers, Washington, DC, 2017.

[3] M. B. Miller, Quantitative Financial Risk Management, Hoboken, NJ: Wiley, 2019.

[4] U.S. Water Resources Council, "Economic and Environmental Principles and Guidelines for Water and Related Land Resources Implementation Studies," U.S. Government Printing Office, Washington, DC, 1983.

[5] U.S. Bureau of Reclamation & U.S. Army Corps of Engineers, "Best Practices in Dam and Levee Safety Risk Analysis," 2019.

[6] Committee on Risk-Based Analysis for Flood Damage Reduction, "Risk Analysis and Uncertainty in Flood Damage Reduction Studies," National Academy Press, Washington, D.C., 2000.

[7] U.S. Army Corps of Engineers, "EM 1110-2-1619 Risk-Based Analysis for Flood Damage Reduction Studies," Washington, D.C., 1996.

[8] D. Vose, Risk Analysis: A Quantitative Guide, West Sussex, England: John Wiley & Sons, Ltd. , 2008.

[9] A. H.-S. Ang and W. H. Tang, Probability Concepts in Engineering Planning and Design: Volume II Decision, Risk, and Reliability, Canada: John Wiley & Sons, Inc., 1984.

[10] U.S. Army Corps of Engineers, "ER 1105-2-100 Planning Guidance Notebook," Washington, D.C., 2000.

[11] U.S. Army Corps of Engineers, "ER 1110-2-1156 Safety of Dams - Policy and Procedures," Washington, D.C., 2014.

[12] C. H. Smith, M. Bartles and M. Fleming, "RMC-TR-2018-03 An Inflow Volume-Based Approach to Estimating Stage-Frequency Curves for Dams," U.S. Army Corps of Enigneers, Lakewood,CO, 2018.

[13] U.S. Army Corps of Engineers, "HEC-FDA Flood Damage Reduction Analysis User's Manual," U.S. Army Corps of Engineers, Davis, CA, 2016.

[14] U.S. Army Corps of Engineers, "ETL 1110-2-537 Uncertainty Estimates for Nonanalytical Frequency Curves," U.S. Army Corps of Engineers, Washington, DC, 1997.

[15] B. Efron and R. J. Tibshirani, An Introduction to the Bootstrap, Boco Raton: CRC Press LLC, 1998.

[16] B. Efron and T. Hastie, Computer Age Statistical Inference: Algorithms, Evidence and Data Science, New York, NY: Cmbridge University Press, 2016.

[17] J. R. M. Hosking and J. R. Wallis, Regional Frequency Analysis: An Approach Based on L-Moments, Cambridge, UK: Cambridge University Press, 1997.

[18] A. R. Rao and K. H. Hamed, Flood Frequency Analysis, Boca Raton, FL: CRC Press LLC, 2000.

[19] K. Krishnamoorthy, Handbook of Statistical Distributions with Applications, Boca Raton, FL: CRC Press, 2016.

[20] J. R. Stedinger and T. A. Cohn, "Flood Frequency Analysis with Historical and Paleoflood Information," *Water Resources Research,* vol. 22, no. 5, pp. 785-793, 1986.

[21] C. H. Smith and B. E. Skahill, "Estimating Design Floods with a Specified Annual Exceedance," in *Australian National Committee on Large Dams (ANCOLD)*, 2019.

[22] C. H. Smith, "RMC-TR-2020-02 Verification of the Bayesian Estimation and Fitting Software (RMC-BestFit)," U.S. Army Corps of Engineers, Lakewood, CO, 2020.

[23] C. H. Smith and M. Doughty, "RMC-TR-2020-03 RMC-BestFit Quick Start Guide," U.S. Army Corps of Engineers, 2020.

[24] A. Stuart and K. Ord, Kendall's Advanced Theory of Statistics: Volume 1 Distribution Theory, 6th ed., vol. 1, New Delhi: Wiley, 1994.

[25] C. H. Smith, "RMC-RFA User's Manual," U.S. Army Corps of Engineers, Lakewood, CO, 2017.

[26] C. H. Smith, "A robust and efficient stochastic simulation framework for estimating reservoir stage-frequency curves with uncertainty bounds," in *Australian National Committee on Large Dams (ANCOLD)*, 2018.

[27] S. Fruhwirth-Schnatter, G. Celeux and C. P. Robert, Handbook of Mixture Analysis, Boca Raton, FL: CRC Press, 2020.

[28] J. R. Stedinger, R. M. Vogel and E. Foufoula-Georgiou, "Chapter 18 Frequency Analysis of Extreme Events," in *Handbook of Hydrology*, McGraw-Hill, 1993, p. 18.40.

[29] U.S. Army Corps of Engineers, "EM 1110-2-1415 Hydrologic Frequency Analysis," Washington, D.C., 1993.

[30] D. N. D. Hartford and G. B. Baecher, Risk and Uncertainty in Dam Safety, London, England: Thomas Telford, 2004.

[31] M. Aboelata, D. S. Bowles and D. M. McClelland, "A Model for Estimating Dam Failure Life Loss," in *Proceedings of the Australian Committee on Large Dams Risk Workshop*, Launceston, Tasmania, Australia, 2003.

[32] U.S. Army Corps of Engineers, "HEC-LifeSim Life Loss Estimation User's Manual," Davis, CA, 2018.

[33] U.S. Army Corps of Engineers, "HEC-LifeSim Technical Reference Manual," Institute for Water Resources, USACE, Davis, CA, 2020.

[34] A. D. Kiureghian, Structural and System Reliability, New York, NY: Cambridge University Press, 2022.

[35] A. Serrano-Lombillo, I. Escuder-Bueno, M. G. de Membrillera-Ortuno and L. Altarejoe-Garcia, "Methodology for the calculation of annualized incremental risks in systems of dams," *Risk Analysis,* 2010.

[36] U.S. Army Corps of Engineers, "Key USACE Flood Risk Management Terms," U.S. Army Corps of Engineers, Davis, CA, 2015.

[37] Australian National Committee on Large Dams (ANCOLD), "Guidelines on Risk Assessment," 2003.

[38] R. E. Moss, Applied Civil Engineering Risk Analysis, Cham, Switzerland: Springer, 2020.

[39] S. G. Vick, Degrees of Belief: Subjective Probability and Engineering Judgement, Reston, VA: American Society of Civil Engineers, 2002.

[40] L. Ruschendorf, Mathematical Risk Analysis: Dependence, Risk Bounds, Optimal Allocations and Portfolios, New York, NY: Springer, 2010.

[41] P. Hill, D. Bowles, P. Jordan and R. Nathan, "Estimating Overall Risk of Dam Failure: Practical Considerations in Combining Failure Probabilities," in *Australian National Committee on Large Dams (ANCOLD) Risk Workshop*, 2003.

[42] P. I. Hill, D. S. Bowles, R. J. Nathan and R. Herveynen, "On the art of event tree modeling for portfolio risk analyses," in *Australian National Committee on Large Dams (ANCOLD)*, Auckland, NZ, 2001.

[43] P. A. Zielinski, "Event Trees in the Assessment of Dam Safety Risks," in *Australian National Committee on Large Dams (ANCOLD)*, Canberra, Australia, 2014.

[44] A. Srivastava, "A Computation Framework for Dam Safety Risk Assessment with Uncertainty Analysis," Utah State University, Logan, Utah, 2013.

[45] R. Jongejan, "RMC-TR-2019-04 Willamette Valley System Risk Analysis," U.S. Army Corps of Engineers, Lakewood, CO, 2019.

[46] L. R. Beard, "Probability Estimates Based on Small Normal-Distribution Samples," *Journal of Geophysical Research,* July 1960.

[47] A. C. Cullen and H. C. Frey, Probabilisitc Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs, New York, NY: Plenum Press, 1999.

[48] Center for Chemical Process Safety, Guidelines for Developing Quantitative Safety Risk Criteria, Hoboken, New Jersey: John Wiley & Sons, Inc., 2009.

[49] R. Rockafellar and S. Uryasev, "Conditional value-at-risk for general loss distributions," *Journal of Banking and Finance,* vol. 26, no. 7, pp. 1443-1471, 2002.

[50] P. Artzner, F. Delbaen, J.-M. Eber and D. Heath, "Coherent measures of risk," *Mathematical Finance,* vol. 9, no. 3, pp. 203-228, 1999.

[51] C. H. Smith, "Improving the Economic Evaluation of Flood Risk Management Studies," Colorado School of Mines, Golden, CO, 2022.

[52] G. B. Baecher and J. T. Christian, Reliability and Statistics in Geotechnical Engineering, West Sussex, UK: John Wiley & Sons Ltd., 2003.

[53] W. M. Bolstad and J. M. Curran, Introduction to Bayesian Statistics, Hoboken, NJ: John Wiley & Sons, Inc., 2017.

[54] Y. Y. Haimes, Risk Modeling, Assessment, and Management, Hoboken, NJ: John Wiley & Sons, Inc., 2004.

[55] R. J. Hyndman and Y. Fan, "Sample Quantiles in Statistical Packages," *The American Statistician,* vol. 50, no. 4, pp. 361-365, 1996.

[56] D. E. Knuth, The Art of Computer Programming, Volume 2: Seminumerical Algorithms, 2nd ed., Reading, MA: Addison-Wesley, 1981.

[57] M. G. Morgan and M. Henrion, Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis, New York, NY: Cambridge University Press, 1990.

[58] A. E. Raftery and S. M. Lewis, "How Many Iterations in the Gibbs Sampler?," *Bayesian Statistics,* vol. 4, pp. 763-773, 1992.

[59] RAC Engineers and Economists, LLC, "Technical Guide for DAMRAE version 4.0.0.0," U.S. Army Corps of Engineers, Lakewood, CO, 2017.

[60] D. P. Amlung, "Spreadsheet Validation: DamonRAE," U.S. Army Corps of Engineers, Lakewood, CO, 2019.

[61] P. J. Davis and P. Rabinowitz, Methods of Numerical Integation, 2nd ed., Mineola, New York: Dover Publications, Inc., 2007.

[62] G. Lepage, "A New Algorithm for Adaptive Multidimensional Integration," *Journal of Computational Physics,* vol. 27, no. 1, pp. 192-203, 1978.

[63] G. Lepage, "VEGAS: An Adaptive Multidimensional Integration Program," Cornell University, 1980.

[64] W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, Numerical Recipes: The Art of Scientific Computing., 3rd ed., Cambridge, UK: Cambridge University Press, 2017.

[65] A. Ciric, A Guide to Monte Carlo & Quantum Monte Carlo Methods, Createspace Independent Publishing Platform, 2016.

[66] J. R. M. Hosking, *lmom: L-Moments,* R package version 2.9, 2022.

[67] A. Viglione, J. R. M. Hosking, F. Laio, A. Miller, E. Gaume, O. Payrastre, J. L. Salinas, C. C. N'guyen and K. Halbert, *nsRFA: Non-Supervised Regional Frequency Analysis,* R package version 0.7-15, 2020.

[68] L. R. Beard, "Technical Report CW-151: Statistical Methods in Hydrology," U.S. Army Corps of Engineers, Sacramento District, Sacramento, CA, 1962.

[69] U.S. Geological Survey, "Guidelines for Determining Flood Flow Frequency Bulletin 17B," 1982.

[70] U.S. Geological Survey, "Guidelines for Determining Flood Flow Frequency Bulletin 17C," https://doi.org/10.3133/tm4B5, 2018.

[71] S. Hubbert, Essential Mathematics for Market Risk Management, West Sussex, UK: John Wiley & Sons, Ltd, 2012.

[72] A. Genz, "Numerical Computation of Multivariate Normal Probabilities," *Journal of Computational and Graphical Statistics,* vol. 1, pp. 141-149, 1992.

[73] M. D. Pandey, "An effective approximation to evaluate multinormal integrals," *Structural Safety,* vol. 20, pp. 51-67, 1998.

[74] X.-X. Yuan and M. D. Pandey, "Analysis of approximations for multinormal integration in system reliability computation," *Structural Safety,* vol. 28, pp. 361-377, 2006.

[75] N.-Z. Sun and A. Sun, Model Calibration and Parameter Estimation for Environmental and Water Resources Systems, New York: Springer, 2010.

[76] I. M. Sobol', "Sensitivity analysis for non-linear mathematical models," *Mathematical Modeling & Computational Experiment,* vol. 1, no. 4, pp. 407-414, 1993.

[77] Office of the Federal Registrar, "Title 44 - Emergency Management and Assistance; Part 65 Identification and Mapping of Special Hazard Areas," Office of the Federal Registrar, Washington, DC, 2022.

[78] U.S. Army Corps of Engineers, "ECB 2019-11 Transition Guidance for Levee System Evaluations for the National Flood Insurance Program (NFIP)," U.S. Army Corps of Engineers, Washington, DC, 2019.

[79] C. H. Smith, W. L. Fields and D. A. Margo, "(DRAFT) RMC-TR-2022-XX Quantitative Risk Analysis with the RMC-TotalRisk Software," U.S. Army Corps of Engineers, Lakewood, CO, 2022.

Appendix A – Probability Fundamentals

This appendix provides an overview of essential probability concepts, axioms, and theorems. A solid grasp of these probability fundamentals is key to understanding how the quantitative risk analysis is performed in RMC-TotalRisk. Similar probability theory overviews are provided in [5], [52], and [53].

Basic Terminology and Concepts

This section provides definitions for basic probability terminology used commonly in probability theory and quantitative risk analyses.

- **Random experiment**: An experiment where the outcome cannot be predicted with certainty. Tossing a coin or rolling a die is an example of a random experiment.
- **Sample space**: The set of all possible outcomes for a single trial of a random experiment. For example, tossing a coin will result in heads or tails.
- **Random variable**: A random variable assigns a value and corresponding probability to each outcome in a sample space. For example, heads = 1 and tails = 0.
- **Event**: An event is a subset of the sample space. Following the coin example, tossing a head would be an event within the sample space.

From any two (or more) events A and B other events can be created using the following operations:

- **Union of two events**: The union of two events A and B denoted $A \cup B$, is the set of outcomes in either A or B (inclusive or).
- **Intersection of two events**: The intersection of two events A and B denoted $A \cap B$ is the set of all outcomes in both A and B.
- **Complement of an event**: The complement of an event A is the set of outcomes not in A, denoted $\bar{A}$.
- **Mutually exclusive events**: If two events have no outcomes in common, they are mutually exclusive events. In that case, the occurrence of A excludes the occurrence of B and vice versa. This is also referred to as *disjoint* events.
- **Collectively exhaustive events**: The set of events is collectively exhaustive if at least one of the events must occur.

Venn diagrams are typically used to illustrate the relationships between events. For example, Figure 51 shows a Venn diagram representing two mutually exclusive events A and B. Figure 52 shows a Venn diagram for two mutually exclusive and collective exhaustive events. For illustration purposes, in Figure 52, A can represent tossing heads on a coin and B represents tossing tails. Notice that the two rectangles are separated from each other and that there is no intersection between them. The coin can only be either heads or tails. Also take notice that the complete sample space Ω is also represented by the diagram. With the coin example, events A and B cover all the events within the entire sample space. This means that events A and B are collectively exhaustive.

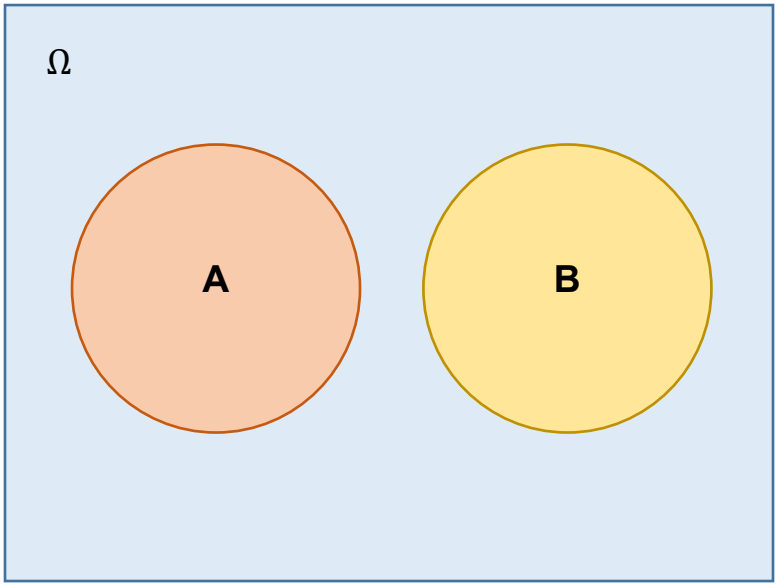


Figure 51 - Venn diagram for two mutually exclusive events.

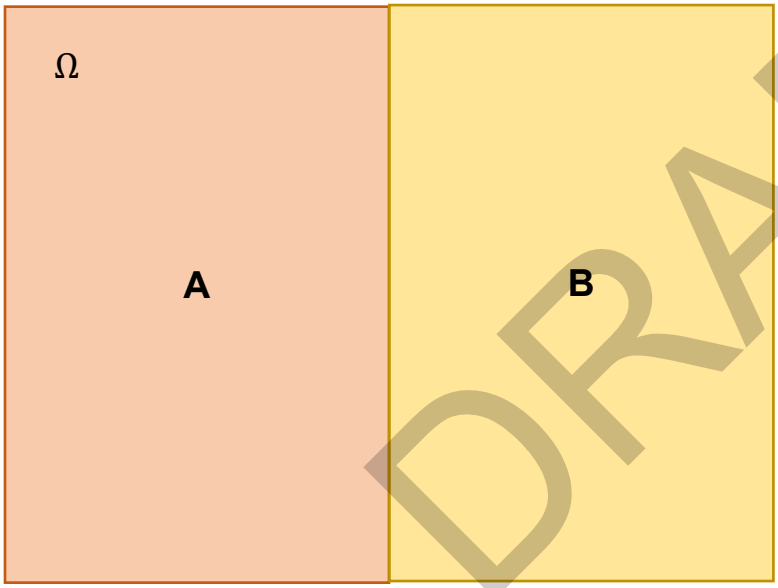


Figure 52 - Venn diagram for two mutually exclusive and collectively exhaustive events.

Figure 53 shows a diagram with intersecting events. The overlapping area is highlighted with a black border to illustrate the intersection between events A and B. An example is picking a random card from a deck of cards. In this case, event A could represent the event of selecting a King, and event B could be the event of a red suit (hearts or diamonds). The intersection event represents selecting a red King.

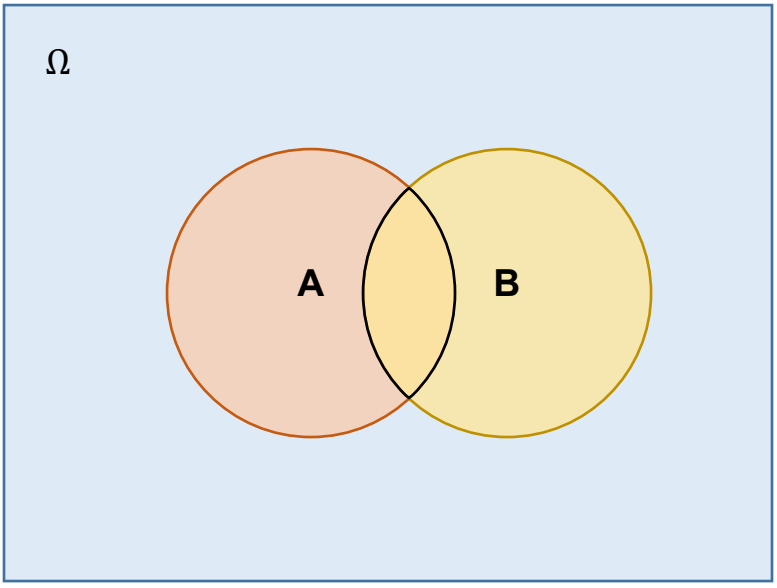


Figure 53 - Venn diagram for intersection of two events.

Finally, Figure 54 shows a diagram for the union of two events. Notice that the black border outlines the full area of the two overlapping events. Visually, the union of events A and B is the area of event A plus the area of event B, minus the overlapping area of A and B. In keeping with the deck of cards example, the union represents the event of selecting a King or a red suited card.

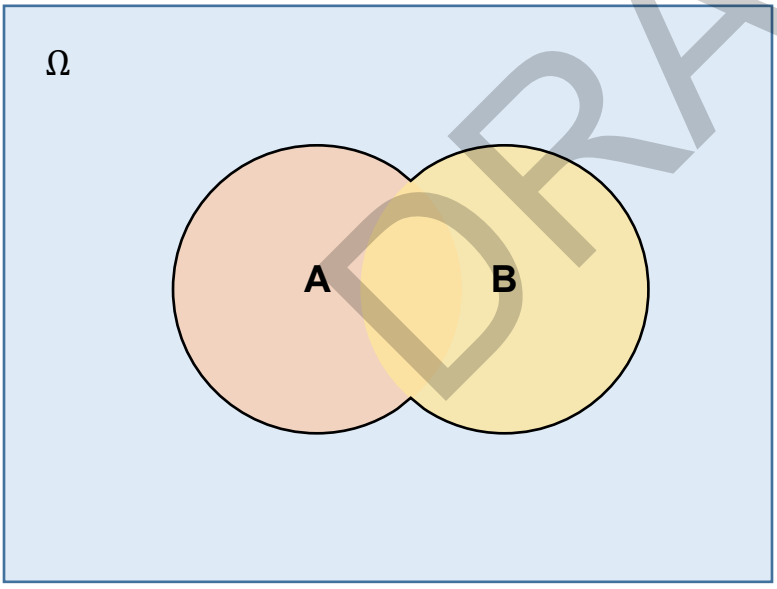


Figure 54 - Venn diagram for the union of two events.

Probability Axioms

Probability can be defined formally by the following axioms:

I. The probability of any event is a non-negative real number.

$$P(A) \geq 0 \qquad \textit{Equation 113}$$

II. The probability that at least one event in the sample space Ω will occur is equal to one.

$$P(\Omega) = 1 \qquad \textit{Equation 114}$$

III. The probability of union of two mutually exclusive events is the sum of their probabilities (the probability is additive over *disjoint* events). This is known as the *addition rule* for probability.

$$P(A \cup B) = P(A) + P(B) \qquad \textit{Equation 115}$$

This axiom can be generalized to represent a set of n mutually exclusive events:

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i) \qquad \textit{Equation 116}$$

Several important rules of probability can be proved from these axioms. This appendix will only discuss the probability rules and theorems most relevant for performing quantitative risk analysis.

Conditional Probability

The probability of an event A, irrespective of the outcome of another variable or event, is referred to as the marginal probability:

$$P(A) \qquad \textit{Equation 117}$$

The conditional probability is the probability of one event occurring in the presence of a second event:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad \textit{Equation 118}$$

An event is statistically independent if the occurrence of another event has no effect on the conditional probabilities of independent events:

$$P(A|B) = P(A) \qquad \textit{Equation 119}$$

For two positively dependent events, an event is more likely to occur in the presence of the other. Therefore, the conditional probability is greater than the marginal:

$$P(A|B) > P(A) \qquad \textit{Equation 120}$$

For two negatively dependent events, an event is less likely to occur in the presence of the other. As such, the conditional probability is less than the marginal:

$$P(A|B) < P(A)$$ *Equation 121*

If two events are mutually exclusive, the occurrence of A excludes the occurrence of B, so the conditional probability is zero:

$$P(A|B) = 0$$ *Equation 122*

Therefore, mutually exclusive events are necessarily negatively dependent, with the only exception being cases where either one or both events have a marginal probability equal to zero. This also means mutually exclusive events are not independent, and independent events cannot be mutually exclusive.

Joint Probability

The joint probability of events A and B is the probability of the two events occurring in the same random experiment. This is also referred to as the probability of intersection:

$$P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$$ *Equation 123*

This is known as the *multiplication rule* or *chain rule* for probability. For independent events, since the conditional probability equals the marginal probability, the joint probability becomes:

$$P(A \cap B) = P(A) \cdot P(B)$$ *Equation 124*

If two events are mutually exclusive, $P(A|B) = 0$, so the joint probability between events A and B is also zero.

Probability of Union

The probability that events A or B (or both) occur is the probability of union of A and B. When the events are mutually exclusive, the probability of union is:

$$P(A \cup B) = P(A) + P(B)$$ *Equation 125*

This is known as the *addition rule* for probability. If the events are not mutually exclusive, the intersection must be subtracted (as seen in Figure 54) as follows:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$ *Equation 126*

If the two events are statistically independent, then the conditional probability equals the marginal probability, so the probability of union becomes:

$$P(A \cup B) = P(A) + P(B) - P(A) \cdot P(B)$$ *Equation 127*

De Morgan's Rule

The probability that event A does not occur is the probability of the complement of event A:

$$P(\bar{A}) = 1 - P(A)$$ *Equation 128*

This is known as the *subtraction rule* for probability. The complement of the union of two events is equal to the intersection of their complements:

$$\overline{A \cup B} = \bar{A} \cap \bar{B} \qquad \textit{Equation 129}$$

This is often referred to as *De Morgan's Law* or *De Morgan's Rule*. De Morgan's Rule can be combined with the probability of a complement to provide an alternative method for calculating the probability of union for two statistically independent events:

$$P(A \cup B) = 1 - [1 - P(A)] \cdot [1 - P(B)] \qquad \textit{Equation 130}$$

This calculation can be generalized to represent a set of n independent events:

$$P\left(\bigcup_{i=1}^{n} E_i\right) = 1 - \prod_{i=1}^{n} [1 - P(E_i)] \qquad \textit{Equation 131}$$

This method simplifies the calculation for the probability of union when there are more than two statistically independent events.

Inclusion-Exclusion Principle

In combinatoric mathematics, the inclusion-exclusion principle generalizes the formula for computing the probability of union for many events with dependency:

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i) - \sum_{1 \le i < j \le n} P(E_i \cap E_j) + \sum_{1 \le i < j < k \le n} P(E_i \cap E_j \cap E_k) - \cdots + (-1)^{n+1} \cdot P(E_i \cap \cdots \cap E_n) \qquad \textit{Equation 132}$$

This can be written in compact notation as:

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{k=1}^{n} (-1)^{k+1} \cdot \left(\sum_{1 \le i_1 < \cdots < i_k \le n} P(E_{i_1} \cap \cdots \cap E_{i_k})\right) \qquad \textit{Equation 133}$$

This equation can be more clearly understood by considering the case of three events A, B, and C (see Figure 55). The probability union is given by:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \qquad \textit{Equation 134}$$

From this equation, the inclusion-exclusion principle can be generalized into the following recursive steps:

1. Include the marginal probability of each event, $P(A) + P(B) + P(C)$
2. Exclude the pairwise joint probability, $-P(A \cap B) - P(A \cap C) - P(B \cap C)$
3. Include the triple-wise joint probability, $+P(A \cap B \cap C)$
4. Continue until the n-tuple-wise joint probability is included (if n is odd) or excluded (if n is even)

Figure 55 shows the Venn diagram for the union of three events. There are 8 possible unique combinations of events. The union $A \cup B \cup C$, which is visualized with the black border, contains 7 of these events.

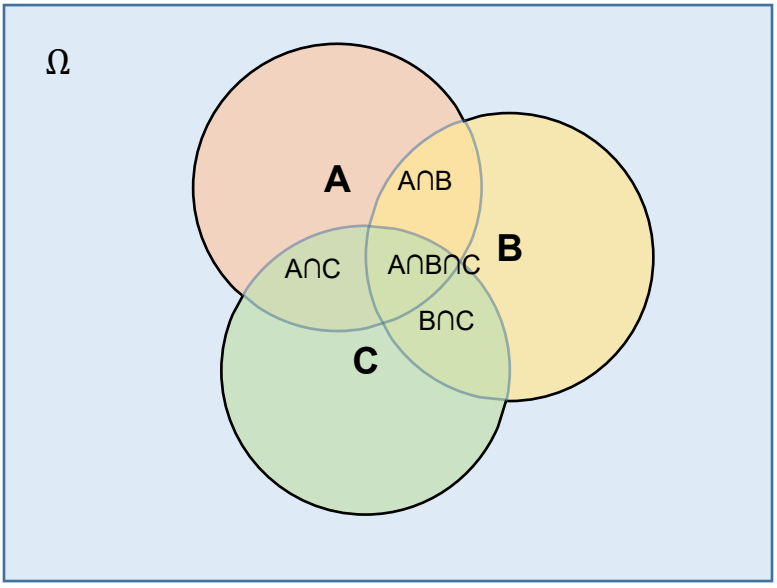


Figure 55 - Venn diagram for the union of three events.

Combinations of Events

Combinatoric mathematics has an important role in how RMC-TotalRisk combines probabilities for multiple failure modes and multiple system elements. Combinatorics is a broad field and there are books on this subject alone. For the sake of brevity, there are two key concepts that are integral for computing system risk. First, is determining the number of ways the system can fail. Failure and non-failure are binary; that is, there are only two possible outcomes for the system. However, if there are multiple failure modes, the number of possible ways the system can fail is:

$$2^n - 1 \qquad \textit{Equation 135}$$

For example, as seen in Figure 55, if there are 3 failure modes (events), there are 7 ways in which the system can fail. As the number of failure modes and system components increase, the number of ways the system can fail increases exponentially.

The second key concept is how to determine the number of ways that k subevents can occur from among n total events when the order of the events does not matter. In other words, the number of k-event combinations in an n-event set. This is given by the binomial coefficient:

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} \qquad \textit{Equation 136}$$

For example, there are 3 combinations of 2-failure event subsets among 3 possible failure modes (events): $P(A \cap B)$, $P(A \cap C)$, and $P(B \cap C)$.

Law of Total Probability

The law of total probability gives the total probability of an event by summing across distinct events. This is also referred to as the *total probability theorem*. The idea behind the law of total probability is illustrated using a Venn diagram in Figure 56. If $B_1, B_2, B_3, \cdots$ is a partition of the sample space Ω, then for any event A, the total probability is given by:

$$P(A) = \sum_i P(A \cap B_i) = \sum_i P(A|B_i) \cdot P(B_i) \qquad \textit{Equation 137}$$

Stated another way, to find the total probability of event A, the sample space Ω can be partitioned into bins. The probability $P(A)$ is then determined by adding the amount of probability of A that falls into each partition. The law of total probability is instrumental in performing quantitative risk analysis.

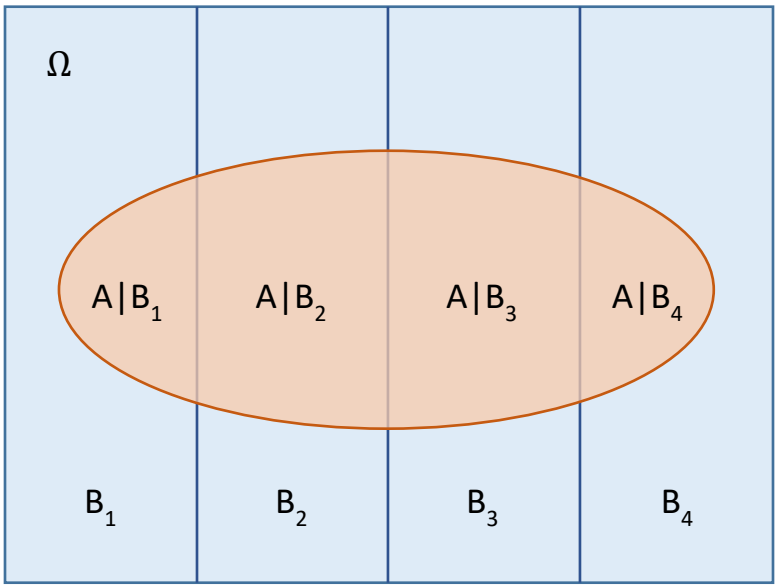


Figure 56 - Venn diagram of the law of total probability.

Bayes' Theorem

Suppose that $P(A|B)$ is known but $P(B|A)$ is not. Using the rules for conditional probability described above, the joint probability of events A and B is:

$$P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A) \qquad \textit{Equation 138}$$

Divide each side by $P(A)$ to obtain:

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)} \qquad \textit{Equation 139}$$

which is Bayes' rule. $P(B)$ is considered the prior probability of B, and $P(B|A)$ is the posterior probability given A. To find $P(A)$ in the denominator, the law of total probability is used as follows:

$$P(B_j|A) = \frac{P(A|B_j) \cdot P(B_j)}{\sum_i P(A|B_i) \cdot P(B_i)} \qquad \textit{Equation 140}$$

While Bayes' theorem is not used directly in RMC-TotalRisk, hazard functions can be imported from the Bayesian Estimation and Fitting software, *RMC-BestFit*. For more details on Bayesian estimation, please see [22] and [23].

Discrete Random Variables

A random variable assigns a value and corresponding probability to each outcome in a sample space. Random variables can be numerical or categorical. A discrete random variable has a countable number of discrete values. For example, there are six sides to a die. The outcome of rolling a die is a discrete random variable. There are exactly six possible values for the random variable. Examples of discrete random variables in risk analysis is the number of spillway gates that can fail or not fail during a flood event. The probability distribution of a discrete random variable is called the probability mass function (PMF):

$$p(x) = P(X = x) \qquad \textit{Equation 141}$$

The cumulative distribution function (CDF) for a discrete random variable is defined as:

$$F(x) = P(X \leq x) \qquad \textit{Equation 142}$$

which is calculated from the sum of the PMF:

$$F(x) = \sum_{t \leq x} p(t) \qquad \textit{Equation 143}$$

The probabilities $P(x)$ must satisfy the following:

$$0 \leq P(x_i) \leq 1 \text{ for each } i \qquad \textit{Equation 144}$$

$$\sum_{x} p(x) = 1 \qquad \textit{Equation 145}$$

Continuous Random Variables

A continuous random variable can be an infinite number of possible values. Examples of continuous random variables in risk analysis are streamflow, water surface elevation, seismic ground motion, and structural resistance to hydraulic loading. The probability for a specific value of a continuous random variable is not defined. Instead, it is defined over an interval (or range) of values. A continuous random variable X has a probability density function (PDF, $f(x)$):

$$P(a \leq X \leq b) = \int_{a}^{b} f(x) \cdot dx \qquad \textit{Equation 146}$$

where a and b are any two numbers, with $a < b$. Then,

$$\int_{-\infty}^{+\infty} f(x) \cdot dx = 1 \qquad \textit{Equation 147}$$

The cumulative distribution function (CDF) can then be defined as:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t) \cdot dt \qquad \textit{Equation 148}$$

where $F(x)$ is the nonexceedance probability ($0 \leq F(x) \leq 1$) for value x of a random variable X. The PDF is the derivative of the CDF:

$$f(x) = \frac{dF(x)}{dx}$$ *Equation 149*

The complementary CDF, commonly referred to as the *survival function*, provides the exceedance probability:

$$S(x) = 1 - F(x) = P(X > x)$$ *Equation 150*

The F-N curves used for portraying risk results are like survival functions except F-N curves are defined as $P(X \geq x)$. Finally, the inverse CDF, commonly called the quantile function, provides the value of x where $P(X \leq x) = p$. The inverse CDF is defined as:

$$F^{-1}(p) = x$$ *Equation 151*

The inverse CDF is instrumental to performing Monte Carlo simulation (see **Appendix C**).

Appendix B – Summary Statistics

This appendix provides mathematic details on the risk result summary statistics provided in RMC-TotalRisk.

Moments

The moments of a distribution provide useful quantitative measures related to the shape of the distribution function. The raw moments (*moments about zero*) of a continuous distribution are defined as:

$$\mu_i' = \int_{-\infty}^{\infty} x^i \cdot f(x) \cdot dx \qquad \textit{Equation 152}$$

and for discrete distributions as:

$$\mu_i' = \sum_{j=0}^{n} x_j^i \cdot p(x_j) \qquad \textit{Equation 153}$$

The zero raw moment is equal to one as required by the second probability axiom:

$$\mu_0' = \int_{-\infty}^{\infty} x^0 \cdot f(x) \cdot dx = \int_{-\infty}^{\infty} f(x) \cdot dx = 1 \qquad \textit{Equation 154}$$

The first raw moment is the mean, or expected value, and it is the center of mass of the distribution:

$$\mu = \mu_1' = \mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f(x) \cdot dx \qquad \textit{Equation 155}$$

The central moments (*moments about the mean*) of a continuous distribution are defined as:

$$\mu_i = \mathbb{E}\left[(X-\mu)^i\right] = \int_{-\infty}^{\infty} (x-\mu)^i \cdot f(x) \cdot dx \qquad \textit{Equation 156}$$

The second central moment is the variance, σ^2, which provides a measure of dispersion about the mean:

$$\sigma^2 = \mu_2 = \mathbb{E}[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x) \cdot dx \qquad \textit{Equation 157}$$

The square root of the variance is the standard deviation, $\sigma = \sqrt{\sigma^2}$. The standardized central moment of a continuous distribution is the central moment normalized by the standard deviation to render the moment scale invariant.

$$\tilde{\mu}_i = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^i\right] = \int_{-\infty}^{\infty} \left(\frac{X-\mu}{\sigma}\right)^i \cdot f(x) \cdot dx \qquad \textit{Equation 158}$$

The third standardized central moment is skewness, which provides a measure of symmetry:

$$\gamma = \tilde{\mu}_3 = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \int_{-\infty}^{\infty}\left(\frac{x-\mu}{\sigma}\right)^3 \cdot f(x) \cdot dx \qquad \textit{Equation 159}$$

Finally, the fourth standardized central moment is kurtosis, which provides a measure of tail thickness or heaviness:

$$\kappa = \tilde{\mu}_4 = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = \int_{-\infty}^{\infty}\left(\frac{x-\mu}{\sigma}\right)^4 \cdot f(x) \cdot dx \qquad \textit{Equation 160}$$

RMC-TotalRisk reports the mean and standard deviation of the F-N curves to provide useful summary statistics related to the shape of those loss distributions.

Conditional Expectation

A conditional expectation is defined as the expected value of a random variable given that this value lies within some prescribed probability range [54]:

$$\mathbb{E}[X|X \geq \beta] = \frac{\int_{\beta}^{\infty} x \cdot f(x) \cdot dx}{\int_{\beta}^{\infty} f(x) \cdot dx} \qquad \textit{Equation 161}$$

The denominator is Equation 161 is the exceedance probability α of the condition β:

$$\alpha = F(x > \beta) = 1 - F(x \leq \beta) \qquad \textit{Equation 162}$$

Substituting Equation 162 into Equation 161, the following is obtained:

$$\mathbb{E}[X|X \geq \beta] = \frac{1}{\alpha}\int_{\beta}^{\infty} x \cdot f(x) \cdot dx \qquad \textit{Equation 163}$$

If $\alpha = 1$ then the conditional expectation $\mathbb{E}[X|X \geq \beta]$ is equivalent to the unconditional expectation $\mathbb{E}[X]$. Equation 163 is commonly referred to as the Conditional Value-at-Risk (CVaR). Alternative names for CVaR found in literature are *Average Value-at-Risk*, *Expected Shortfall*, and *Expected Tail Loss*.

The conditional expectation depends on where the probability axis of the distribution is partitioned. The distribution can be partitioned by conditioning on x values or probability values. In economics and finance, CVaR is often estimated based on a specified exceedance probability α rather than a threshold hazard level or event, β. In this case, $\beta = F^{-1}(1-\alpha)$.

In RMC-TotalRisk, a conditional expectation is computed for each risk type. For the risk of failure and incremental risk, the conditional expectation is conditional on the critical hazard level x_c beyond which there is a nonzero probability of failure $P(F|x_c) > 0$. For example, the conditional expectation for incremental risk is:

$$\eta_{\Delta} = \mathbb{E}[C_{\Delta}|x \geq x_c] = \frac{\sum_{i=x_c}^{n} P(x_i) \cdot P(F|x_i) \cdot C_{\Delta}(x_i)}{\sum_{i=x_c}^{n} P(x_i) \cdot P(F|x_i)} = \frac{\mathbb{E}[C_{\Delta}]}{\alpha_{\Delta}} \qquad \textit{Equation 164}$$

where $C_{\Delta}(x_i) = C_F(x_i) - C_{NF}(x_i)$; x_c is the threshold hazard level; and α_{Δ} is the exceedance probability for which incremental consequences would occur.

Percentiles

In statistics, given an array of data $\boldsymbol{x} = \{x_1, x_2, \dots, x_n\}$, the k-th percentile gives the nonparametric inverse CDF, $F^{-1}(k) = x_k$. If the k-th percentile is 0.9 (90%), then 90% of the values in $\boldsymbol{x}$ are less than $F^{-1}(0.9) = x_{0.9}$. There is no standard method for computing a percentile, but all methods yield similar results when the number of observations is very large. RMC-TotalRisk uses Definition 7 from Hyndman and Fan [55] which is the same method used by the PERCENTILE and PERCENTILE.INC functions in Microsoft Excel®. First, the data values are sorted in ascending order. Then, the location of the k-th percentile is computed as:

$$m = (n-1) \cdot k + 1 \qquad \textit{Equation 165}$$

where n is the length of the value array $\boldsymbol{x} = \{x_1, x_2, \dots, x_n\}$. If $m = 1$ then return the first value in the array x_1. If $m = n$ then return the last value in the array x_n. Otherwise, the k-th percentile x_k is interpolated as:

$$i = floor(m) \qquad \textit{Equation 166}$$

$$x_k = x_i + (m - i) \cdot (x_{i+1} - x_i) \qquad \textit{Equation 167}$$

Kernel Density Estimation

To visualize the distribution of an array of Monte Carlo output data $\boldsymbol{x} = \{x_1, x_2, \dots, x_n\}$, RMC-TotalRisk performs kernel density estimation (KDE). KDE is a nonparametric method for estimating the probability density function of a random variable. Let $\boldsymbol{x} = \{x_1, x_2, \dots, x_n\}$ be independent and identically distributed samples. The kernel density estimator is estimated as:

$$f(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \qquad \textit{Equation 168}$$

where n is the length of the value array $\boldsymbol{x} = \{x_1, x_2, \dots, x_n\}$; h is the bandwidth; and K is the kernal. TotalRisk uses the standard normal distribution for the kernel function, $K = \phi$. The bandwidth is estimated as:

$$h = \sigma \left(\frac{4}{3n}\right)^{\frac{1}{5}} \qquad \textit{Equation 169}$$

Where σ is the sample standard deviation of $\boldsymbol{x}$. To compute percentiles, the data is sorted in ascending order, and given a Weibull plotting position $p_i = \frac{r}{n+1}$, where r is the rank of the i-th data value. Then, the nonparametric inverse CDF is computed from linear interpolation:

$$F^{-1}(p) = x_i + (x_{i+1} - x_i)\left(\frac{p - p_i}{p_{i+1} - p_i}\right) \qquad \textit{Equation 170}$$

Figure 57 provides an example of the kernel density plot and summary statistics provided for risk measures.

Figure 57 - Example of the kernel density and percentile output for risk measures.

Histogram

A histogram is an approximate representation of the distribution of an array of data $\boldsymbol{x} = \{x_1, x_2, \dots, x_n\}$. A histogram can be thought of as a simplified kernel density estimation. Histograms are constructed by first binning the range of values and then counting how many values fall into each bin. By default, RMC-TotalRisk uses the "Rice Rule" to select the number of bins:

$$k = 2\sqrt[3]{n} \qquad \textit{Equation 171}$$

Then, the bin width is estimated as:

$$h = \frac{\max \boldsymbol{x} - \min \boldsymbol{x}}{k} \qquad \textit{Equation 172}$$

The bins are then specified as consecutive, non-overlapping intervals of equal size. Figure 58 shows an example of the histogram and summary statistics output that is displayed when right-clicking the table column header. From Figure 57 and Figure 58, the kernel density and histogram plots both provide similar information on the distribution of the output data.

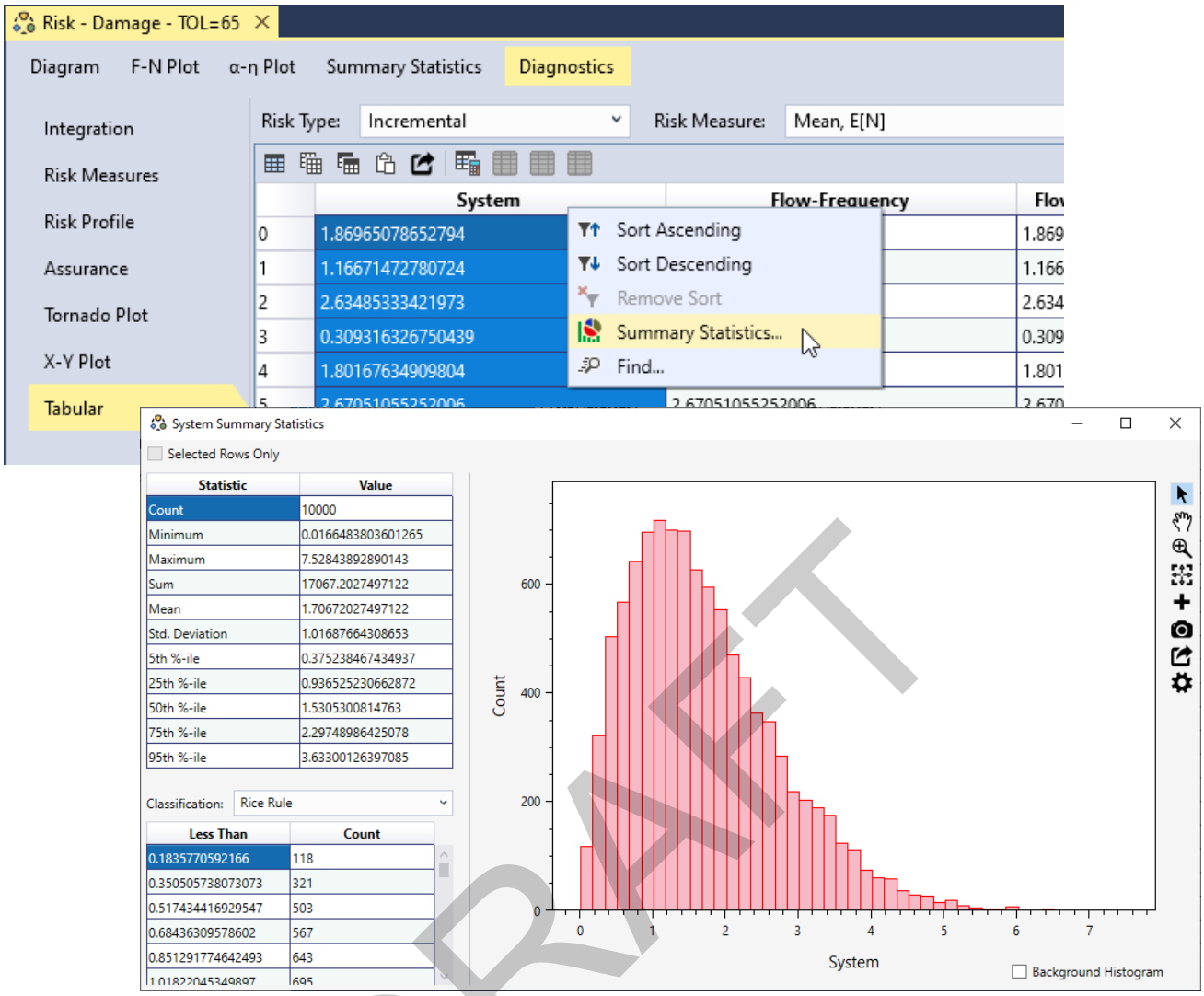


Figure 58 - Example of the histogram and summary statistics output for tabular data.

Appendix C – Monte Carlo Simulation

Monte Carlo simulation uses repeated sampling to solve a problem numerically. Monte Carlo simulation is typically used for three types of numerical problems: optimization, integration, and uncertainty analysis. RMC-TotalRisk uses Monte Carlo to perform uncertainty analysis.

The first step in a Monte Carlo analysis is to construct a model, such as an event tree or risk analysis in RMC-TotalRisk. The second step is to define probability distributions that describe the uncertainty in model inputs, such as the uncertainty in the hazard, system response, and consequence functions. Finally, the Monte Carlo simulation is performed, where a value is drawn at random from the distribution of each model input. Together, this set of random values, one for each input, defines a scenario, which is used as input to the model to compute corresponding model output values. The entire process is repeated N times producing N independent scenarios with corresponding output values. Each set of random values, or scenarios, is commonly referred to as a *realization*. The N realizations for each model output can then be used to quantify the uncertainty in the model outputs.

Inverse Transform Sampling

Inverse transform sampling is a method commonly used to generate the random samples of a random variable from its distribution. The method is derived from the inverse probability integral transform, which states that if Y has a uniform distribution on the closed interval [0,1] and X is a random variable having a cumulative distribution function of $F_X(X)$, then the random variable $F_X^{-1}(Y)$ has the same distribution as X.

A key aspect of a Monte Carlo simulation is the random number generator. It is important to understand that these random numbers are generated from a deterministic algorithm and are more correctly called pseudorandom numbers. A pseudorandom number generator (PRNG) is used to create uniform random numbers in the interval $[0, 1]$. A PRNG seed initializes the sequence of random numbers and can be set to ensure repeatability of the Monte Carlo results. All RMC software uses the subtractive PRNG algorithm built into the Microsoft® .NET framework [56].

A random number generated by the PRNG corresponds to the CDF (nonexceedance probability) of the input probability distribution. For example, consider the distribution of an uncertain input variable X. As shown in Appendix A, the CDF gives the probability that the variable X is less than or equal to x:

$$F(x) = P(X \leq x) \qquad \textit{Equation 173}$$

$F(x)$ ranges from 0 to 1. The inverse CDF describes the value of x for a given nonexceedance probability, p:

$$F^{-1}(p) = x \qquad \textit{Equation 174}$$

The inverse CDF (or quantile function) is central to the generation of random samples from each input distribution in a Monte Carlo simulation. To generate a random sample for a probability distribution, a random number $r \sim U[0, 1]$ is generated and then input into the inverse CDF to calculate the corresponding x value that is randomly generated from the distribution:

$$x = F^{-1}(r) \qquad \textit{Equation 175}$$

For demonstration, assume an input has a Normal distribution with a mean of 100 and standard deviation of 15. Figure 59 shows the inverse CDF of the distribution and provides a graphical representation of the Monte Carlo sampling process. It shows that the random number of $r = 0.748$ corresponds to a value of $x = 110$.

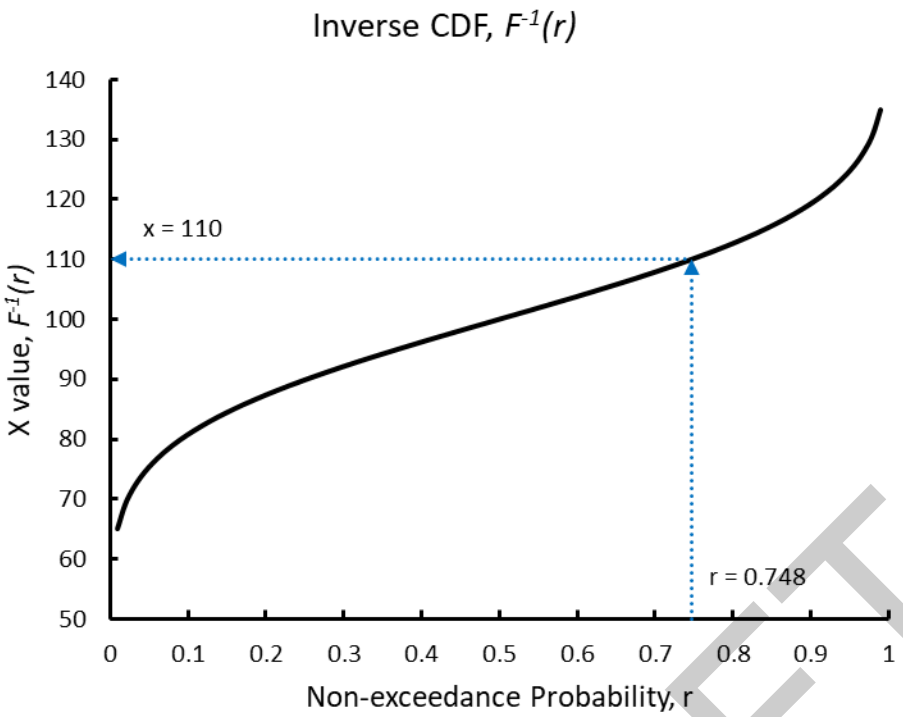


Figure 59 – Illustration of Monte Carlo sampling process with the inverse CDF.

Convergence

In a Monte Carlo simulation, many random numbers are generated and the corresponding x values are recorded. Convergence is guaranteed as the number of samples approaches infinity. However, Monte Carlo simulation results can exhibit substantial statistical fluctuation for small samples. The choice in the number of Monte Carlo realizations N should depend on both the simulation runtime, and the intended use of the results. If precision in the mean is the most important consideration, the minimum number of realizations N can be estimated as follows:

$$N_{min} = \left\{ \frac{\sigma}{\varepsilon \cdot \mu} \cdot \Phi^{-1}\left(\frac{1+\alpha}{2}\right) \right\}^2 \qquad \textit{Equation 176}$$

where μ is the output mean; σ is the output standard deviation; ε is the desired error (e.g., $\varepsilon = 0.01 = 1\%$ error); $\Phi^{-1}(\cdot)$ is the inverse CDF of the standard Normal distribution; and α is the confidence interval width. For example, with a mean of 100 and standard deviation of 15, a minimum of 865 realizations are required to be 95% confident the Monte Carlo output mean is $100 \pm 1\%$.

Another criterion for selecting the number of Monte Carlo realizations is the precision of a percentile (fractile) of the output, e.g., the 5th and 95th percentiles define the 90% confidence interval of a model output. In this case, the minimum number of realizations is calculated as:

$$N_{min} = p \cdot (1-p) \cdot \left\{ \frac{\Phi^{-1}\left(\frac{1+\alpha}{2}\right)}{\Delta p} \right\}^2 \qquad \textit{Equation 177}$$

where p is the percentile of interest (e.g., 0.95); Δp is the tolerance for the percentile (e.g., ± 0.01); $\Phi^{-1}(\cdot)$ is the inverse CDF of the standard Normal distribution; and α is the confidence interval width. For example, a minimum of

$1{,}825$ realizations are required to be 95% confident the Monte Carlo output 95th percentile is between the true 94th and 96th percentiles $[p - \Delta p, p + \Delta p]$.

To further illustrate the concept of convergence, Figure 60 shows a Monte Carlo simulation with just 100 realizations. The histogram (shown in red) of the 100 samples does not closely match the true input distribution (shown in black). Figure 61 shows a Monte Carlo simulation with $1{,}000$ realizations. The simulation results are beginning to converge towards the true input distribution. Finally, Figure 62 shows a Monte Carlo with $10{,}000$ realizations. From this plot, it is clear the simulation results have closely converged towards the true distribution. The percent errors in the output mean and 95th percentile is provided in Table 8 and Table 9, respectively. Figure 63 and Figure 64 illustrate the Monte Carlo convergence in the output mean and 95th percentile as the number of realizations increases towards $10{,}000$.

In RMC-TotalRisk, the default number of realizations in the risk analysis is $10{,}000$. This ensures the risk analysis runtime is manageable and provides sufficient precision for the output mean and confidence interval percentiles. More information on selecting the number of Monte Carlo realizations is provided in [57] and [58].

Table 8 - Assessing Monte Carlo convergence in the output mean.

Realizations	Simulated Mean	True Mean	% Error
100	97.69	100	2.31%
1,000	99.75	100	0.25%
10,000	99.97	100	0.03%

Table 9 - Assessing Monte Carlo convergence in the output 95th percentile.

Realizations	Simulated 0.95	True 0.95	% Error
100	122.94	124.67	1.39%
1,000	125.28	124.67	0.49%
10,000	124.87	124.67	0.16%

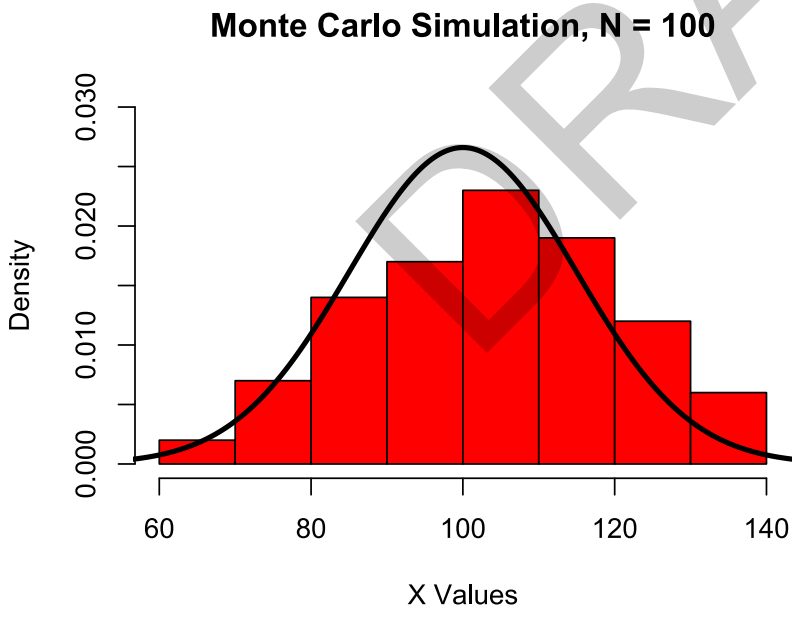


Figure 60 - Monte Carlo Simulation with 100 samples.

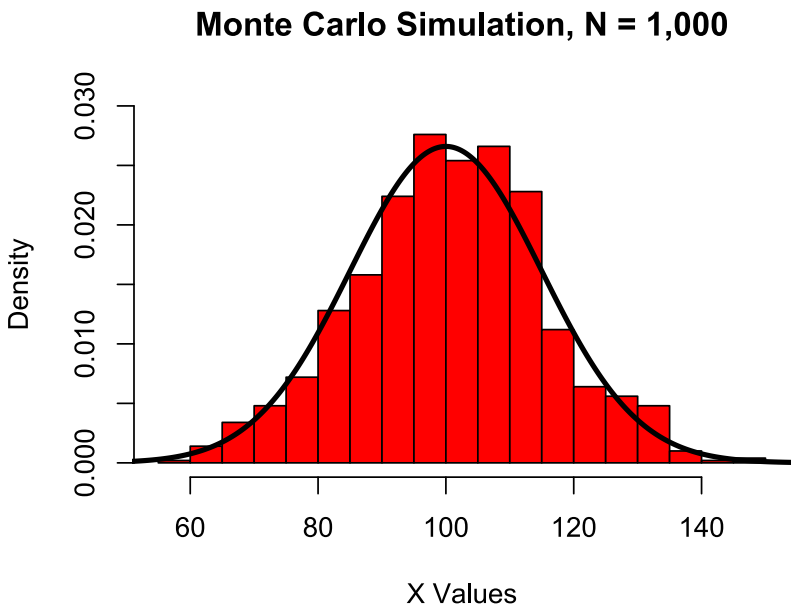


Figure 61 - Monte Carlo Simulation with 1,000 samples.

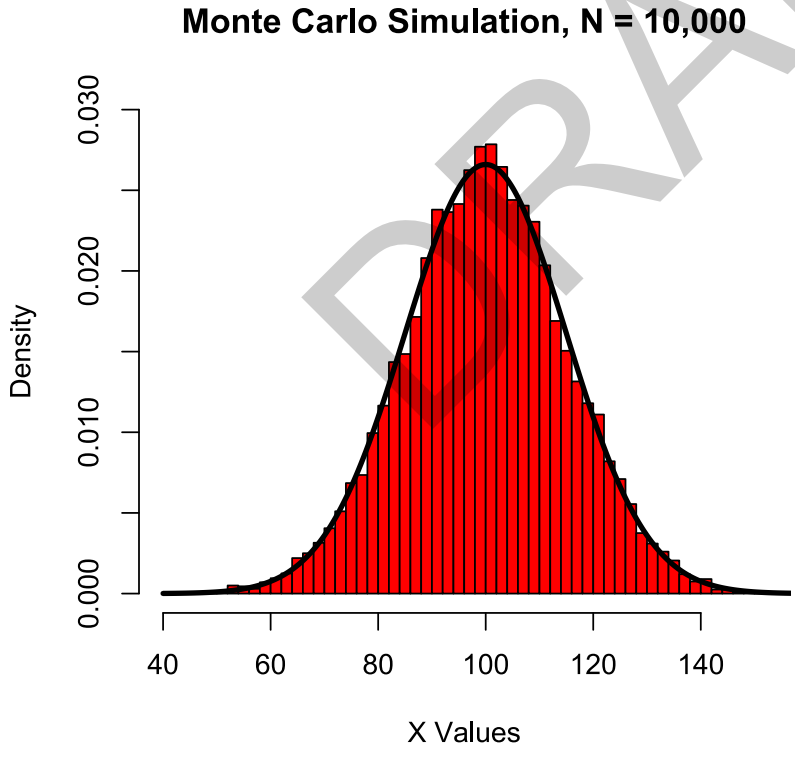


Figure 62 - Monte Carlo Simulation with 10,000 samples.

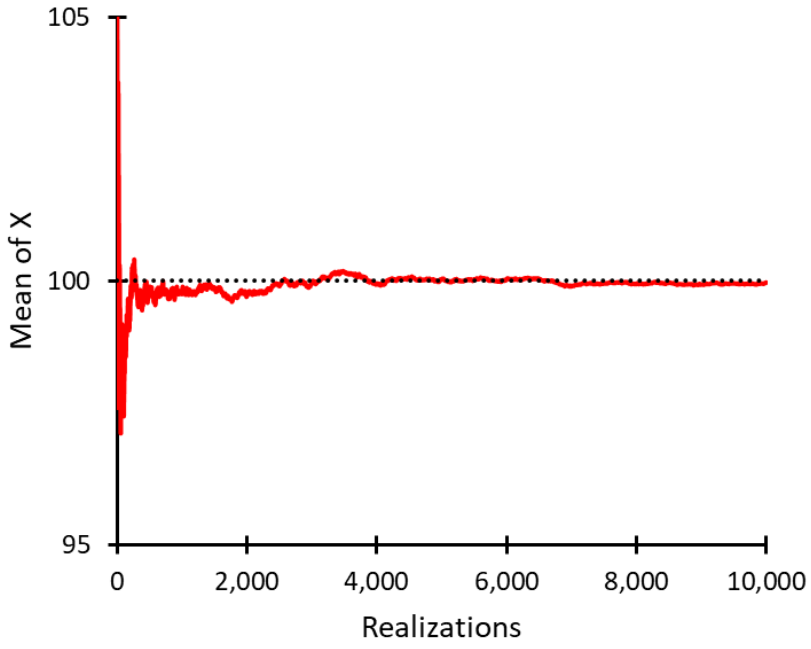


Figure 63 – Illustration of Monte Carlo convergence in the output mean.

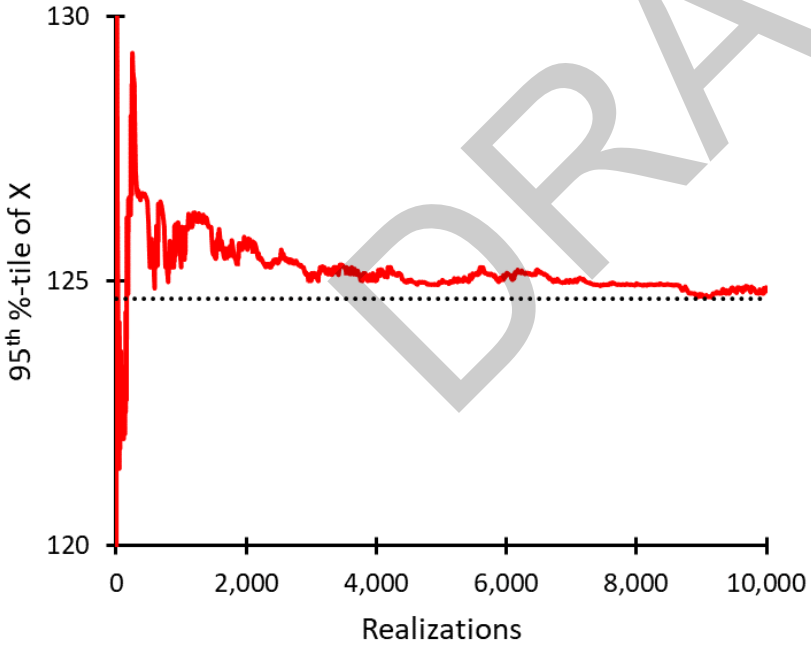


Figure 64 – Illustration of Monte Calo convergence in the output 95th percentile.

Appendix D – Numerical Integration

In RMC-TotalRisk, within every Monte Carlo realization, risk is computed using numerical integration. To establish a foundation in numerical integration techniques, this appendix provides a tutorial on integrating a probability distribution using the trapezoidal rule. Then, a simple risk analysis example is provided to demonstrate the steps required to compute the risk of failure.

Next, more sophisticated adaptive integration techniques are introduced and demonstrated. RMC-TotalRisk uses two adaptive integration methods: 1) Adaptive Simpson's Rule is used for single component system risk analyses; and 2) VEGAS, an adaptive importance sampling method, is used for multiple component system risk analyses. Algorithmic details on the two adaptive integration methods used in RMC-TotalRisk are provided.

Integrating a Probability Distribution

Recall that risk is generally computed as the expected value of a probability distribution:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f(x) \cdot dx \qquad \textit{Equation 178}$$

where $f(x)$ is the PDF of the probability distribution. Definite integrals can be numerically solved with Riemann sums, such as the trapezoidal rule. The method works by approximating the region under the function $g(x) = x \cdot f(x)$ as a trapezoid and calculating its area:

$$\mathbb{E}[X] = \int_{a}^{b} g(x) \cdot dx \approx \left(\frac{g(a) + g(b)}{2}\right) \cdot (b - a) \qquad \textit{Equation 179}$$

$$\mathbb{E}[X] = \int_{a}^{b} x \cdot f(x) \cdot dx \approx \left(\frac{a \cdot f(a) + b \cdot f(b)}{2}\right) \cdot (b - a) \qquad \textit{Equation 180}$$

This approximation can be improved by partitioning (or binning) the integration interval $[a, b]$ and then applying the trapezoidal rule to each subinterval and summing the results:

$$\mathbb{E}[X] = \sum_{i=1}^{K} \left(\frac{x_{a_i} \cdot f(x_{a_i}) + x_{b_i} \cdot f(x_{b_i})}{2}\right) \cdot (x_{b_i} - x_{a_i}) \qquad \textit{Equation 181}$$

To ensure the integration of the probability distribution is collectively exhaustive $\int_a^b f(x) \cdot dx = 1$, rectangular areas can be added to the first and last subintervals as follows:

$$\mathbb{E}[X] = x_{a_1} \cdot F(x_{a_1}) + \sum_{i=1}^{K} \left(\frac{x_{a_i} \cdot f(x_{a_i}) + x_{b_i} \cdot f(x_{b_i})}{2}\right) \cdot (x_{b_i} - x_{a_i}) + x_{b_K} \cdot \{1 - F(x_{b_K})\} \qquad \textit{Equation 182}$$

where $F(\cdot)$ is the CDF; K is the number of integration bins (or subintervals); and the total number of computational steps is $K + 2$. The trapezoidal rule accuracy improves with increasing number of bins, K.

Example calculations of the trapezoidal rule integration of a distribution using Equation 182 is provided in Table 10 below. In this example, the goal is to compute the mean of a Gamma distribution with a scale of $\theta = 5$ and shape of $\kappa = 10$. The true mean is $\mu = 50$. The integration limits were set as $a = F^{-1}(0.001) = 14.80$ and $b = F^{-1}(0.999) = 113.29$. There are five integration bins, with a total of $5 + 2 = 7$ computation steps. The mean from numerical integration is estimated as 49.981, which has less than a 0.1% error from the true value.

The five trapezoidal integration bins are show in Figure 65 shown in black dots. The true distribution is shown as solid blue. The five bins are enough to capture the general shape of the true distribution. The accuracy will improve by increasing the number of bins.

Table 10 - Example calculations of the trapezoid rule integration of a distribution using the PDF.

Bin	x_a	x_b	$\overline{x \cdot f(x)}$	dx	$\overline{x \cdot f(x)} \cdot dx$	
					0.015	$\leftarrow x_{a_1} \cdot F(x_{a_1})$
1	14.80	34.50	0.343	19.70	6.763	
2	34.50	54.20	0.945	19.70	18.607	
3	54.20	73.89	0.866	19.70	17.065	
4	73.89	93.59	0.315	19.70	6.213	
5	93.59	113.29	0.061	19.70	1.204	
					0.113	$\leftarrow x_{b_5} \cdot \{1 - F(x_{b_5})\}$
				$\mathbb{E}[X] = \sum =$	**49.981**	

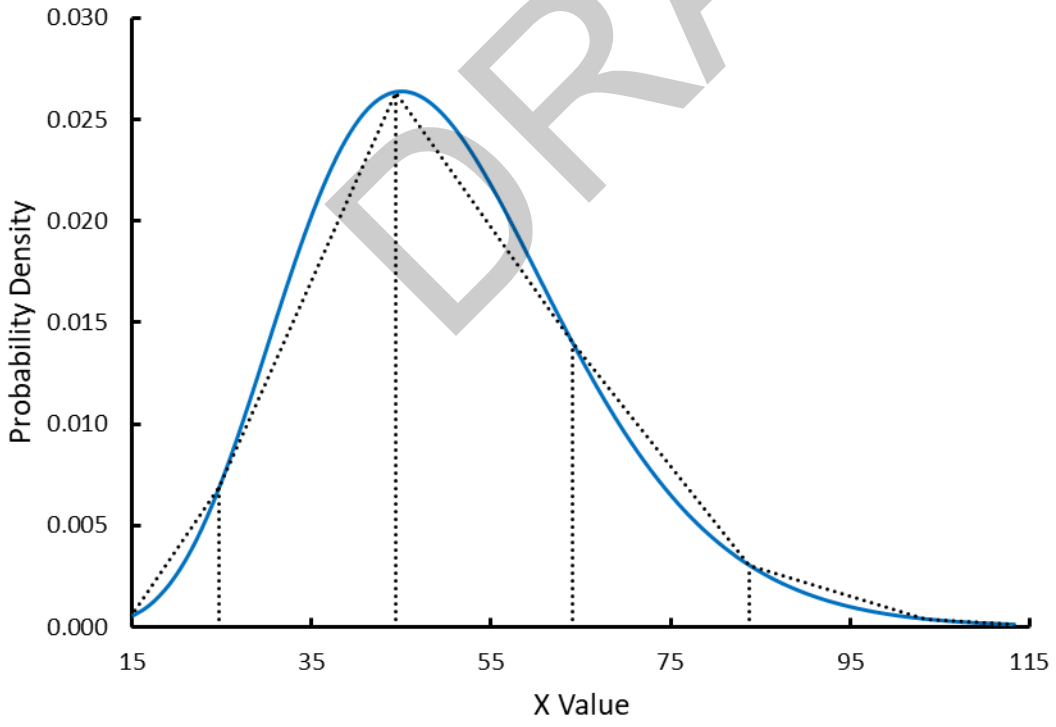


Figure 65 - Example of trapezoidal bins versus the Gamma PDF.

An alternative and often more accurate way to integrate a probability distribution is to use the CDF. As shown in **Appendix A**, the PDF $f(x)$ is the derivative of the CDF:

$$f(x) = \frac{dF(x)}{dx} = \frac{F(x_u) - F(x_l)}{x_u - x_l} \qquad \textit{Equation 183}$$

Plugging this into Equation 178 yields:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot dF(x) \qquad \textit{Equation 184}$$

$$\mathbb{E}[X] = \int_{a}^{b} x \cdot dF(x) \approx \left(\frac{a+b}{2}\right) \cdot \{F(b) - F(a)\} \qquad \textit{Equation 185}$$

Then, after partitioning the integration interval and ensuring $\int_a^b dF(x) = 1$, the integral is calculated as:

$$\mathbb{E}[X] = x_{a_1} \cdot F(x_{a_1}) + \sum_{i=1}^{K} \left(\frac{x_{a_1} + x_{b_i}}{2}\right) \cdot \{F(x_{b_i}) - F(x_{a_1})\} + x_{b_K} \cdot \{1 - F(x_{b_K})\} \qquad \textit{Equation 186}$$

where K is the number of integration bins (or subintervals); and the total number of computational steps is $K + 2$.

Example calculations are provided in Table 11. The mean is estimated as 49.99, which has less than a 0.1% error from the true value. As before, the accuracy will improve by increasing the number of bins. Figure 66 below shows that the limited number of trapezoidal bins closely matches the shape of the Gamma CDF, but there is still some error in the approximation.

Table 11 - Example calculations of the trapezoid rule integration of a distribution using the CDF.

Bin	x_a	x_b	$\bar{x}$	dF	$\bar{x} \cdot dF$	
				0.001	0.015	$\leftarrow x_{a_1} \cdot F(x_{a_1})$
1	14.80	34.50	24.65	0.158	3.907	
2	34.50	54.20	44.35	0.482	21.389	
3	54.20	73.89	64.04	0.281	17.987	
4	73.89	93.59	83.74	0.067	5.609	
5	93.59	113.29	103.44	0.009	0.969	
				0.001	0.113	$\leftarrow x_{b_5} \cdot \{1 - F(x_{b_5})\}$
				$\mathbb{E}[X] = \sum =$	**49.990**	

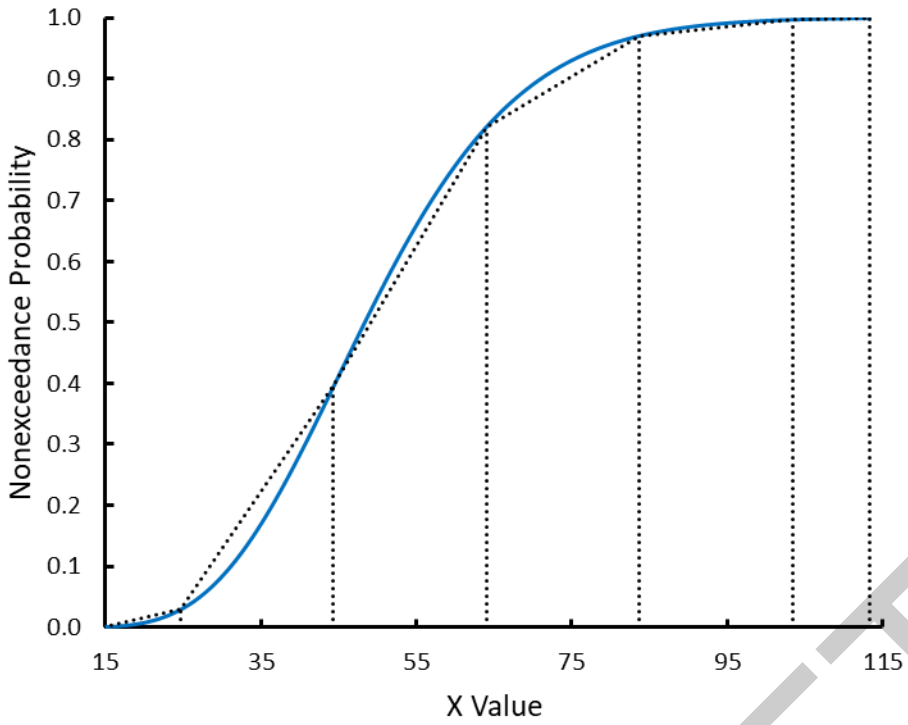


Figure 66 - Example of trapezoidal bins versus the Gamma CDF.

The final way to integrate a probability distribution is to use the inverse CDF:

$$\mathbb{E}[X] = \int_0^1 F^{-1}(p) \cdot dp \qquad \textit{Equation 187}$$

where $F^{-1}(\cdot)$ is the inverse CDF, $F^{-1}(p) = x$; and $dp = 1$. In the cases where a distribution is unbounded, values close to 0 and 1 should be used. For example, 10^{-16} and $1 - 10^{-16}$ are sufficiently close to 0 and 1, respectively.

$$\mathbb{E}[X] = \int_a^b F^{-1}(p) \cdot dp \approx \left(\frac{F^{-1}(a) + F^{-1}(b)}{2}\right) \cdot (b - a) \qquad \textit{Equation 188}$$

Then, after partitioning the integration interval and ensuring $\int_a^b dp = 1$, the integral is calculated as:

$$\mathbb{E}[X] = F^{-1}(p_{a_1}) \cdot p_{a_1} + \sum_{i=1}^{K} \left(\frac{F^{-1}(p_{a_i}) + F^{-1}(p_{b_i})}{2}\right) \cdot (p_{b_i} - p_{a_i}) + F^{-1}(p_{b_K}) \cdot \{1 - p_{b_K}\} \qquad \textit{Equation 189}$$

where $F(\cdot)$ is the CDF; K is the number of integration bins (or subintervals); and the total number of computational steps is $K + 2$.

Example calculations are provided in Table 12 below. The mean is estimated as 52.02, which has greater than a 4% error from the true value. Figure 67 below shows that the trapezoidal bins do not match the inverse CDF well, especially on the left and righthand tails. This is because the probability axis of a distribution can be highly nonlinear.

Table 12 - Example calculations of the trapezoid rule integration of a distribution using the inverse CDF.

Bin	p_a	p_b	$\bar{x}$	dp	$\bar{x} \cdot dp$	
				0.0010	0.015	$\leftarrow F^{-1}(p_{a_1}) \cdot p_{a_1}$
1	0.001	0.201	25.638	0.1996	3.907	
2	0.201	0.400	40.502	0.1996	21.389	
3	0.400	0.600	48.450	0.1996	17.987	
4	0.600	0.799	57.462	0.1996	5.609	
5	0.799	0.999	87.921	0.1996	0.969	
				0.0010	0.113	$\leftarrow F^{-1}(p_{b_5}) \cdot \{1 - p_{b_5}\}$
				$\mathbb{E}[X] = \sum =$	**52.019**	

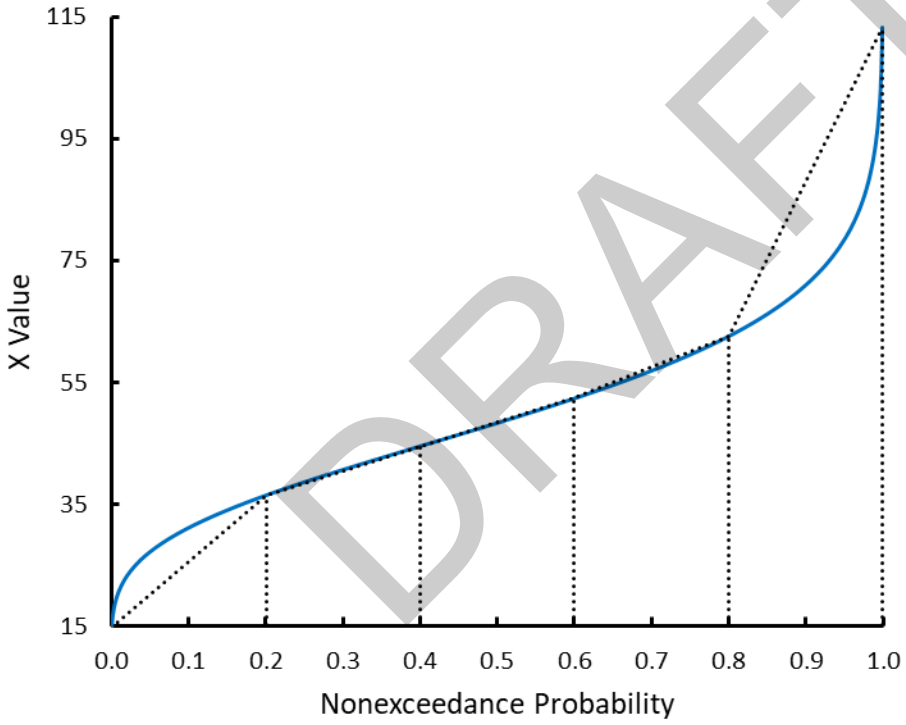


Figure 67 - Example of trapezoidal bins versus the Gamma inverse CDF.

This nonlinearity can be handled by first transforming the probabilities to be approximately linear. Then, stratify the transformed values into K bins and convert back to probability space. Table 13 below provides example calculations where the probabilities were transformed to be Normal z variates, stratified into five bins, then converted back to probabilities. The mean is estimated as 50.83, which has less than a 2% error from the true value. Figure 68 below shows that the trapezoidal bins now more closely match the inverse CDF. As with the previous examples, the accuracy of the numerical integration will improve by increasing the number of bins.

Table 13 - Example calculations of the trapezoid rule integration of a distribution using the inverse CDF with Normal z-transformed probabilities.

Bin	p_a	p_b	$\overline{x}$	dp	$\overline{x} \cdot dp$	
				0.0010	0.015	$\leftarrow F^{-1}(p_{a_1}) \cdot p_{a_1}$
1	0.0010	0.0319	19.903	0.0309	0.614	
2	0.0319	0.2683	32.195	0.2364	7.611	
3	0.2683	0.7317	48.978	0.4635	22.699	
4	0.7317	0.9681	70.818	0.2364	16.742	
5	0.9681	0.9990	98.176	0.0309	3.030	
				0.0010	0.113	$\leftarrow F^{-1}(p_{b_5}) \cdot \{1 - p_{b_5}\}$
				$\mathbb{E}[X] = \sum =$	**50.825**	

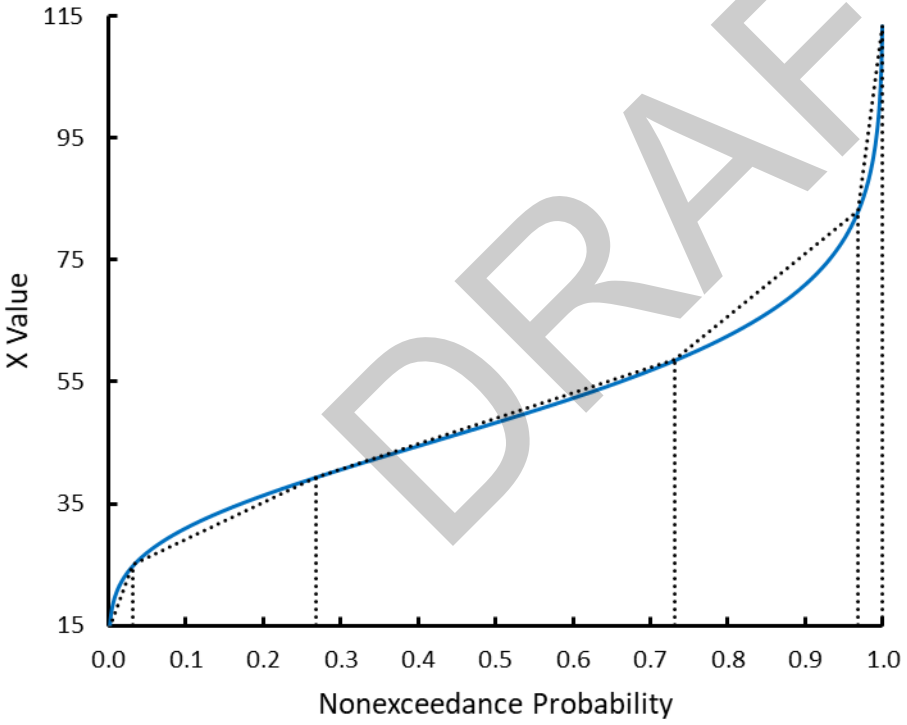


Figure 68 - Example of trapezoidal bins versus the Gamma inverse CDF with Normal z-transformed probabilities.

Risk Analysis Example

The previous examples provide a foundation in numerical integration techniques for risk analysis. Most existing risk analysis software for USACE dam and levee safety or flood risk management studies compute expected annual consequences in one of the three ways presented above.

For example, USACE supported the development of the Dam Safety Risk Analysis Engine, *DAMRAE*©, which is a commonly used software tool for performing event tree calculations and risk analysis for dam safety risk assessment studies [59]. *DAMRAE*© uses a similar integration approach as the second example where numerical integration is performed using the trapezoidal rule with the CDF of the hazard distribution. In recent years, the RMC has used a *Palisade's @Risk*© spreadsheet tool colloquially referred to as *DamonRAE*, which is named after the developer of the tool [60]. *DamonRAE* follows the second approach as well. Likewise, the Flood Damage Reduction Analysis software, *HEC-FDA*, integrates damage exceedance curves following the second example where numerical integration is performed using the trapezoidal rule with the CDF.

On the other hand, the USACE Levee Screening Tool (LST) follows the third approach where integration using the inverse CDF of the hazard distribution. RMC-TotalRisk also follows the third approach using the inverse CDF of the hazard distribution. The primary reason for using the inverse CDF approach in RMC-TotalRisk is that dependency between hazards can be accounted for during the integration process using the Multivariate Normal distribution (see **Appendix F**). The inverse CDF approach can have larger approximation error due to the nonlinearity of the probability axis, as shown above in Figure 67 and Figure 68. However, in TotalRisk, these limitations are overcome by using more sophisticated adaptive integration approaches, which will be discussed in the next section.

In this section, a simple risk analysis example is provided to demonstrate the steps required to compute the risk of failure using the CDF integration approach. This example is designed to show how risk has typically been computed in the past, and to highlight some of the limitations of this approach.

This example risk analysis is for a hypothetical dam, where the hazard function is defined by a Ln-Normal distribution with a mean $\mu = 85$ and standard deviation $\sigma = 20$. In this example, the hazard is the annual max peak reservoir stage (ft). There is a single system response function, which is defined by a Normal distribution with a mean $\mu = 160$ and standard deviation $\sigma = 15$. The response function describes the probability of failure given the reservoir stage. Finally, the consequences (lives lost) of failure are provided as a tabular function in Table 14.

Table 14 - Example risk analysis consequences of failure.

Stage(ft)	Lives
60	0
100	0
140	10
200	100
250	150

In most risk software, it is typical for the full range of the hazard function to be stratified for integration. As such, in this example the integration limits are set at the 0.999 annual exceedance probability (AEP) and the $1e^{-6}$ AEP, which translates to reservoir stage limits of $a = 40.4$ and $b = 249.4$ feet, respectively. For this example, only 20 integration bins were evaluated.

Figure 69 shows the inputs functions with the hazard and response distributions are plotted as CDFs. Figure 70 shows them plotted as PDFs. From these two figures, it can be seen that non-zero probabilities of failure and non-zero consequences do not begin to occur until around a reservoir stage of 100 ft. Figure 71 and Figure 72 show the same plots with the 20 integration bins overlayed. The integral bins begin at $a = 40.4$. This means that the first six integrations bins do not contribute to the overall risk integral. This type of uninformed and static binning was typical in previous USACE risk analysis software. Nearly a quarter of the integration effort is expended on parts of the function that do not contribute to the risk.

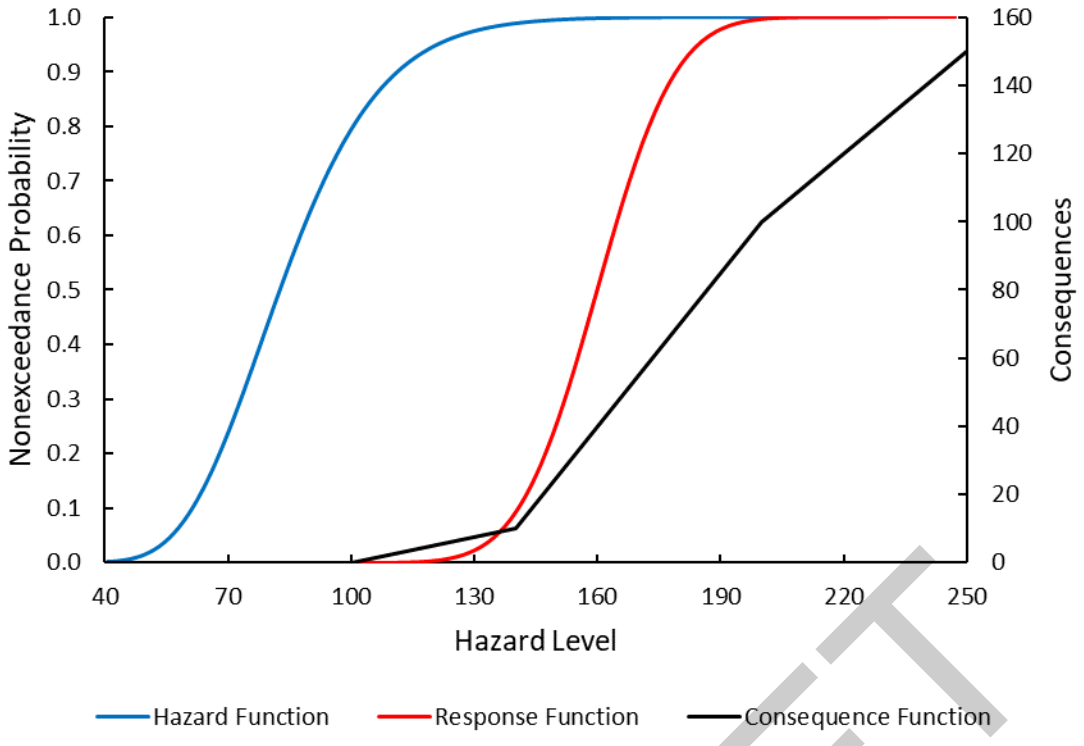


Figure 69 - Example risk analysis input functions where hazard and response functions are plotted as CDFs.

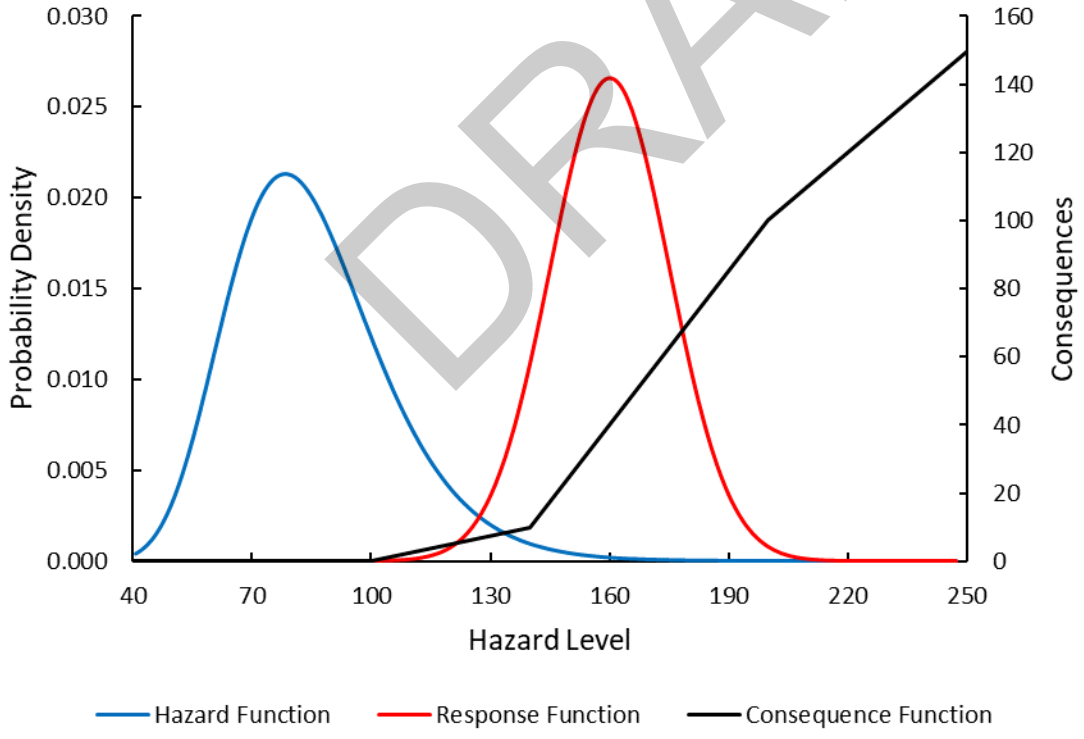


Figure 70 - Example risk analysis input functions where hazard and response functions are plotted as PDFs.

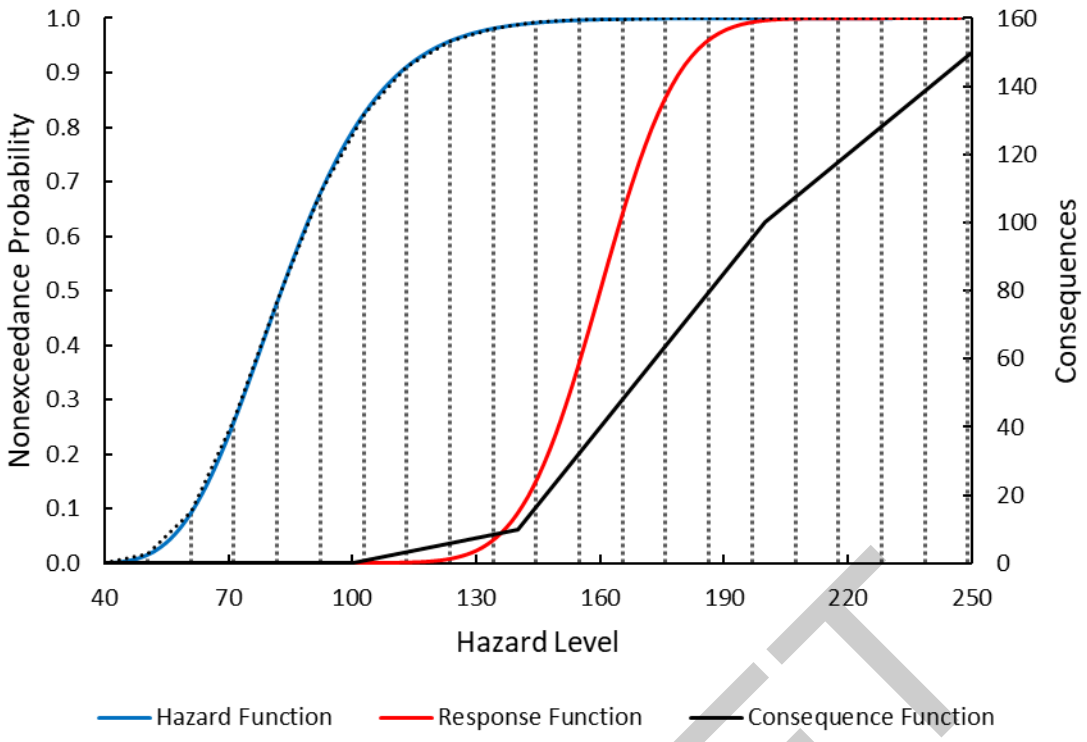


Figure 71 - Example risk analysis input functions where hazard and response functions are plotted as CDFs with 20 trapezoidal integration bins.

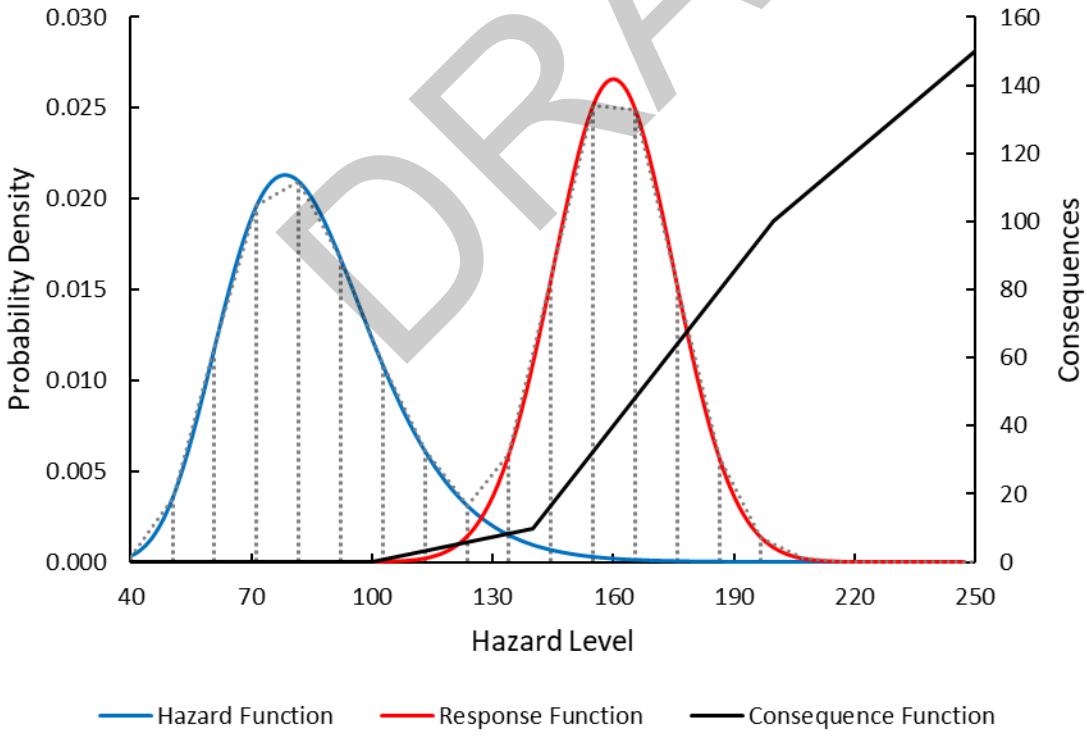


Figure 72 - Example risk analysis input functions where hazard and response functions are plotted as PDFs with 20 trapezoidal integration bins.

As shown in the **Quantitative Risk Analysis** chapter, the risk of failure is computed as follows:

$$\mathbb{E}[C_F] = \int_a^b dF_S(x) \cdot F_{R|S}(x) \cdot C_F(x) \approx \sum_i P(x_i) \cdot P(F|x_i) \cdot C_F(x_i) \qquad \textit{Equation 190}$$

where $P(x_i)$ is the probability of the hazard level x_i; $P(F|x_i)$ is the conditional probability of failure given the hazard level x_i.; and $C_F(x_i)$ is the consequence of failure given the hazard level x_i. The risk calculations following the trapezoidal rule shown in Equation 186 are provided in Table 15 below.

Table 15 - Example calculations of the trapezoid rule integration of a distribution using the CDF.

Bin	x_a	x_b	$\bar{x}$	$dF(x) = P(x)$	$P(F\|\bar{x})$	$C_F(\bar{x})$	$P(x) \cdot P(F\|\bar{x}) \cdot C_F(\bar{x})$
				1.000E-03	1.209E-14	0.00	0.000E+00
1	40.38	50.83	45.61	1.692E-02	1.209E-14	0.00	0.000E+00
2	50.83	61.28	56.06	8.006E-02	2.113E-12	0.00	0.000E+00
3	61.28	71.74	66.51	1.714E-01	2.294E-10	0.00	0.000E+00
4	71.74	82.19	76.96	2.191E-01	1.549E-08	0.00	0.000E+00
5	82.19	92.64	87.41	1.983E-01	6.524E-07	0.00	0.000E+00
6	92.64	103.09	97.87	1.415E-01	1.720E-05	0.00	0.000E+00
7	103.09	113.54	108.32	8.534E-02	2.850E-04	2.08	5.059E-05
8	113.54	124.00	118.77	4.569E-02	2.992E-03	4.69	6.415E-04
9	124.00	134.45	129.22	2.244E-02	2.009E-02	7.31	3.294E-03
10	134.45	144.90	139.67	1.036E-02	8.769E-02	9.92	9.008E-03
11	144.90	155.35	150.13	4.568E-03	2.552E-01	25.19	2.936E-02
12	155.35	165.80	160.58	1.950E-03	5.153E-01	40.87	4.107E-02
13	165.80	176.25	171.03	8.134E-04	7.689E-01	56.54	3.537E-02
14	176.25	186.71	181.48	3.340E-04	9.239E-01	72.22	2.229E-02
15	186.71	197.16	191.93	1.357E-04	9.834E-01	87.90	1.173E-02
16	197.16	207.61	202.38	5.481E-05	9.976E-01	102.38	5.599E-03
17	207.61	218.06	212.84	2.207E-05	9.998E-01	112.84	2.490E-03
18	218.06	228.51	223.29	8.887E-06	1.000E+00	123.29	1.096E-03
19	228.51	238.97	233.74	3.584E-06	1.000E+00	133.74	4.793E-04
20	238.97	249.42	244.19	1.450E-06	1.000E+00	144.19	2.091E-04
				1.000E-06	1.000E+00	144.19	1.442E-04
						$\mathbb{E}[C_F] = \sum =$	**0.163**

Adaptive Integration

The challenge with static numerical integration methods, such as the trapezoidal rule presented above, is that the user must specify the limits of integration and the number of integration bins. If the integrand function has subregions with high variance, then this approach can result in large approximation errors. In addition, as the previous risk analysis example demonstrated, many integrand functions have significant weight in only a narrow subregion. Therefore, many of the integration bins can be wasted in areas of the function that do not contribute significant weight.

Adaptive integration is a type of numerical integration method that adaptively refines the subintervals within the bounds of integration. Adaptive integration methods seek to concentrate the subintervals in the regions that make the largest contribution to the integral, overcoming the limitations of static approaches.

RMC-TotalRisk uses two different adaptive integration approaches. The Adaptive Simpson's Rule is used for single dimension integrations. The VEGAS algorithm, which is an adaptive importance sampling method, is used for multidimensional integrals. Both methods are described in greater detail in the following sections.

Adaptive Simpson's Rule

In RMC-TotalRisk, single dimension integrals are solved using an implementation of the Adaptive Simpson's Rule (ASR) method. The ASR algorithm subdivides the interval of integration in a recursive manner until a user-defined tolerance is achieved. The default tolerance level is 10^{-8}. In each subinterval, the Simpson's rule is used to approximate the region under the function $g(x) = x \cdot f(x)$ as a weighted average of the trapezoidal and midpoint method:

$$\mathbb{E}[X] = \int_a^b g(x) \cdot dx \approx \left[g(a) + 4 \cdot g\left(\frac{a+b}{2}\right) + g(b) \right] \cdot \frac{(b-a)}{6} \qquad \textit{Equation 191}$$

The inputs required for this method are as follows:

- **Max Evaluations**: The maximum number of integrand evaluations allowed when performing numerical integration. This value must be between $10{,}000$ and $1{,}000{,}000$. The default value is $1{,}000{,}000$.
- **Max Depth**: The maximum recursion depth when performing ASR integration. The default is 100.
- **Tolerance**: The desired tolerance for the adaptive Simpson's rule integration. The default is 10^{-8}.

The criterion for determining when to stop subdividing an interval is:

$$\frac{1}{15} \cdot |S(a,m) + S(m,b) - S(a,b)| \leq \epsilon + \epsilon \cdot |S(a,b)| \qquad \textit{Equation 192}$$

where $[a, b]$ is the integration interval; $m = \frac{a+b}{2}$; $S(\cdot)$ is given by the Simpson's rule evaluated at those intervals; and ϵ is the error tolerance for the interval. Each subinterval is recursively subdivided and evaluated until the tolerance is achieved. For more details on the ASR method are provided in [61].

Using the previous risk analysis example, the ASR method requires 333 function evaluations to reach a tolerance of 10^{-8}. Figure 73 and Figure 74 show risk function plots with the ASR integration steps overlayed. The ASR method automatically concentrates the subintervals in the region that is contributing the most to risk, beginning at a hazard level of 100. The ASR method only expends minimal effort in the regions that do not contribute. By using the ASR method, RMC-TotalRisk can ensure a higher level of precision in the risk results with little input from the user required.

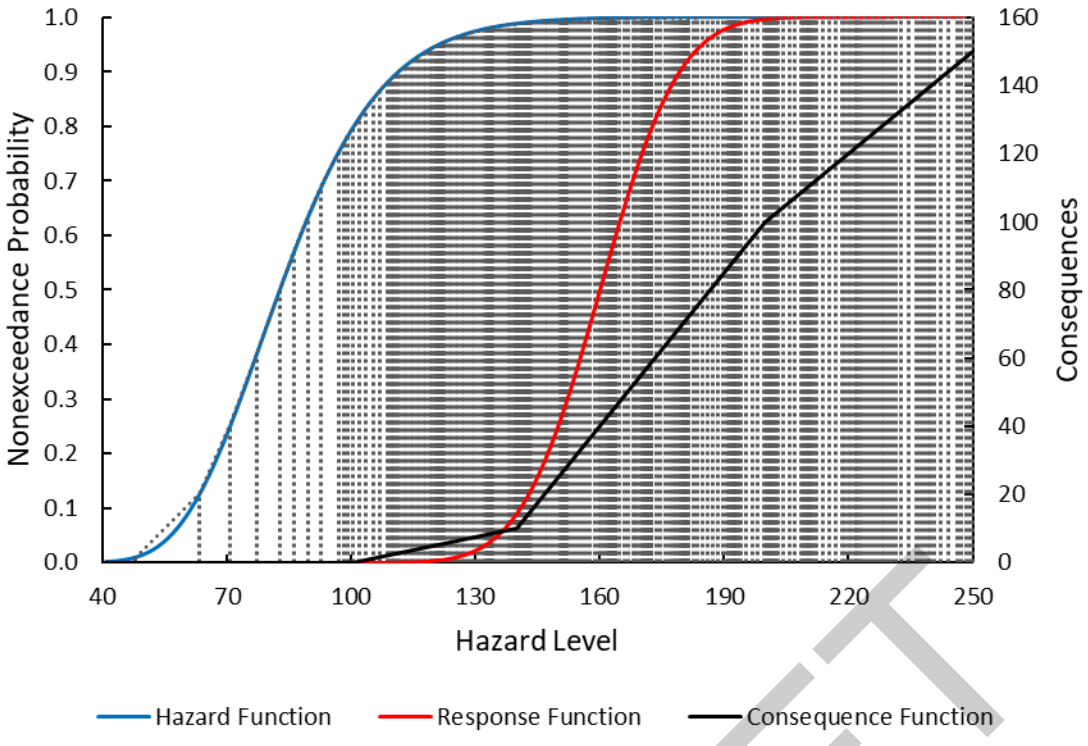


Figure 73 - Example risk analysis input functions where hazard and response functions are plotted as CDFs with the ASR integration bins.

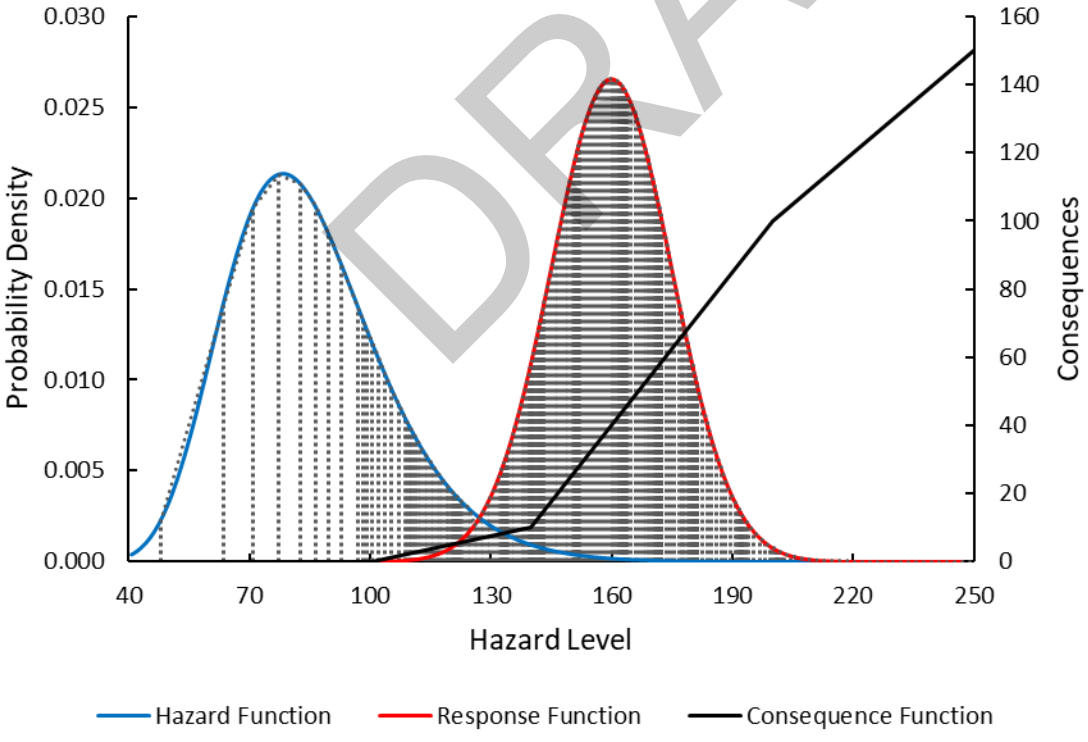


Figure 74 - Example risk analysis input functions where hazard and response functions are plotted as PDFs with the ASR integration bins.

Adaptive Importance Sampling

Computing risk for multiple system components requires integration over a multidimensional integral. Consider a system with two components, where the consequences of failure from each component are additive. Following the general risk formula, the system risk becomes a two-dimensional integral:

$$\mathbb{E}[C]_{\Omega} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{C_X(x) + C_Y(y)\}\cdot f_{XY}\big(C_X(x), C_Y(y)\big)\cdot dx\cdot dy \qquad \textit{Equation 193}$$

where x is the hazard level for system component X; $C_X(x)$ determines the consequences for the hazard level x; y is the hazard level for system component Y; $C_Y(y)$ determines the consequences for the hazard level y; and $f_{XY}\big(C_X(x), C_Y(y)\big)$ is the joint PDF of the combined system consequences occurring.

Solving multidimensional integrals is computationally demanding. If traditional, nonadaptive numerical integration techniques were used, the solution would require K^D iterations, where K is the number of integration steps (or bins) and D is the number of dimensions. For example, if there were 100 integration steps and 5 dimensions, the solution would need 10 billion iterations.

To avoid these computational limitations, RMC-TotalRisk uses an adaptive importance sampling algorithm called VEGAS [62] [63]. This approach is like the ASR method previously described in that evaluation points are concentrated in the regions that make the largest contribution to the integral.

Simple Monte Carlo integration can suffer from low efficiency. As the previous risk analysis example demonstrated, many functions of interest have significant weight in only a few regions. Importance sampling overcomes this problem by increasing the density of evaluation points in the key regions of interest.

In simple Monte Carlo integration, the integral is calculated as an arithmetic mean from N samples:

$$\mathbb{E}[X] = \mu = \int_{-\infty}^{\infty} g(x)\cdot dx \approx \frac{1}{N}\sum_{i=1}^{N} g(x_i) \qquad \textit{Equation 194}$$

where $g(\cdot)$ is the integrand function; and N is the number of Monte Carlo samples. The variance of the mean μ is computed as:

$$\sigma_{\mu}^2 = \frac{1}{N}\sum_{i=1}^{N}(g(x_i) - \mu)^2 \qquad \textit{Equation 195}$$

With importance sampling, the integral is calculated as a weighted mean:

$$\mathbb{E}[X] = \mu = \int_{-\infty}^{\infty} g(x)\cdot p(x)\cdot dx = \int_{-\infty}^{\infty}\left(g(x)\cdot\frac{p(x)}{q(x)}\right)\cdot q(x)\cdot dx \approx \frac{1}{N}\sum_{i=0}^{N} g(x_i)\cdot\frac{p(x_i)}{q(x_i)} \qquad \textit{Equation 196}$$

where $w(x) = \frac{p(x)}{q(x)}$ is the importance weight function, which leads to:

$$\mathbb{E}[X] = \mu_I = \int_{-\infty}^{\infty} g(x)\cdot dx \approx \frac{1}{N}\sum_{i=0}^{N} g(x_i)\cdot w(x_i) \qquad \textit{Equation 197}$$

The variance of the mean μ_I is then:

$$\sigma_{\mu_I}^2 = \frac{1}{N}\sum_{i=1}^{N}(g(x_i)\cdot w(x_i) - \mu_I)^2 < \sigma_{\mu}^2 = \frac{1}{N}\sum_{i=1}^{N}(g(x_i) - \mu)^2 \qquad \textit{Equation 198}$$

Since the variance of the estimate is reduced, the importance sampling approach is a more efficient than basic Monte Carlo integration. Accurate results can be achieved with far fewer function evaluations. More details on importance sampling can be found in [34], [64], and [65].

The VEGAS algorithm approximates the exact risk distribution by making several warmup cycles over the integration region while histogramming the multidimensional integrand function $g_D(x_1, x_2 \cdots, x_D)$. After each cycle, the histogram bins are adapted and refined, and then used to define the sampling distribution for the next cycle. Asymptotically, this procedure will converge to the desired distribution. To avoid the number of histogram bins growing to K^D, the probability distribution is approximated by a separable function $g(x_1)\cdot g(x_2)\cdot \cdots \cdot g(x_D)$ so that the number of histogram bins required is only $K \cdot D$. This algorithm is most efficient when the peaks of the integrand are well localized, which will often be the case for most dam and levee risk analyses.

The algorithm implemented in RMC-TotalRisk has two steps. First, to establish and refine the importance sampling histogram, by default the routine performs five warmup cycles, each with a maximum of $1{,}000 \cdot D$ integrand evaluations. For instance, if $D = 2$, by default there are $10{,}000$ total warmup evaluations. Next, again by default, $10{,}000$ final integrand evaluations are performed. The solution and resulting standard error are based only on the final evaluations. The user can add more function evaluations to achieve a smaller standard error. The following options can be adjusted by the user:

- **Max Evaluations**: The maximum number of integrand evaluations allowed when performing numerical integration. This value must be between $10{,}000$ and $1{,}000{,}000$. The default value is $1{,}000{,}000$.
- **Warmup Cycles**: The number of warmup cycles for the adaptive importance sampling method. The default is 5 warmup cycles.
- **Warmup Evaluations**: The number of integrand evaluations each warmup cycle. The default is $1{,}000 \cdot D$ warmup evaluations, where D is the number of integrand dimensions (i.e., system components).
- **Final Evaluations**: The final number of integrand evaluations for the adaptive importance sampling method. The default is $10{,}000$ final integrand evaluations.

.

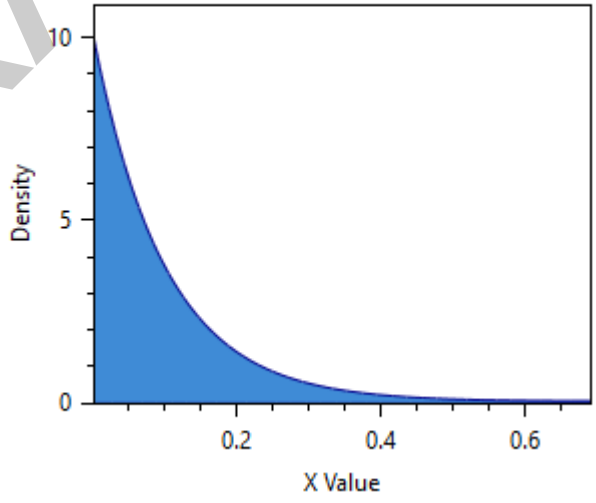
Appendix E – Univariate Probability Distributions

This appendix provides background information and mathematic details of the input probability distributions included in RMC-TotalRisk.

Exponential

The two-parameter (or shifted) Exponential distribution is a special case of the Gamma family of distributions, which includes the two-parameter Gamma, Pearson Type III, and Log-Pearson Type III. The Exponential distribution describes the time between events in a Poisson process, i.e., a process in which events occur continuously and independently at a constant average rate. The Exponential distribution is often used in reliability applications, where it can be used to model data with a constant failure rate. This distribution is useful for modeling highly positively skewed data that have a non-zero lower bound.

Table 16 - Summary table for the Exponential distribution.

Parameters	**Support**
Location: ξ Scale: α	$\begin{cases} -\infty < \xi < +\infty \\ 0 < \alpha < +\infty \end{cases}$
Distribution Functions	**Domain**
$z = \frac{x - \xi}{\alpha}$ $f(x) = \frac{1}{\alpha} e^{-(z)}$ $F(x) = 1 - e^{-(z)}$ $F^{-1}(p) = \xi - \alpha \ln(1 - p)$	$\xi \leq x < +\infty$ Location (ξ) 0 Scale (α) 0.1 

Gamma

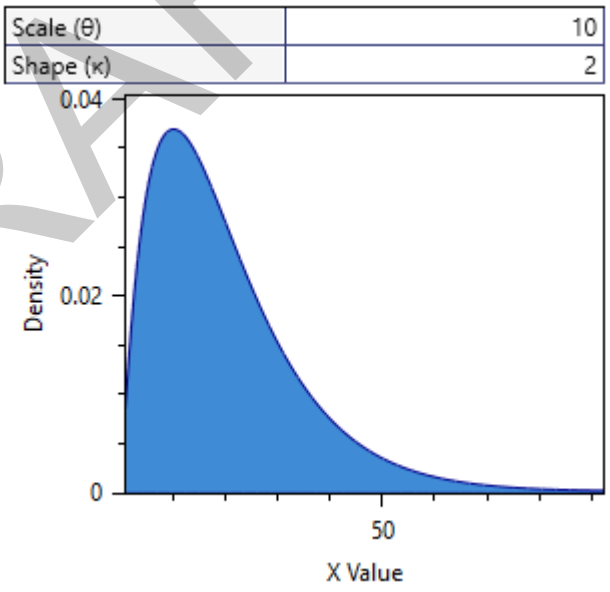
The Gamma distribution is a two-parameter, positively skewed distribution. The Exponential, Erlang, and Chi-squared distributions are special cases of the Gamma distribution. The Gamma distribution has applications in hydrology, econometrics, and other applied fields. It is a flexible distribution capable of modeling many kinds of data with a mild positive skew and lower bound of zero.

There are three different parameterizations for the Gamma distribution in common use:

1. With a scale parameter θ and a shape parameter κ.
2. With an inverse scale parameter $\beta = 1/\theta$, called a rate parameter, and a shape parameter $\alpha = \kappa$.
3. With a mean parameter $\mu = k\theta = \alpha/\beta$ and a shape parameter κ.

RMC software uses the first parameterization, with a scale parameter θ and a shape parameter κ.

Table 17 - Summary table for the Gamma distribution.

Parameters	**Support**
Scale: θ Shape: κ	$\begin{cases} 0 < \theta < +\infty \\ 0 < \kappa < +\infty \end{cases}$
Distribution Functions	**Domain**
$f(x) = \frac{1}{\Gamma(\kappa)\theta^{\kappa}} x^{\kappa-1} e^{-\frac{x}{\theta}}$ where $\Gamma(\cdot)$ is the gamma function $F(x) = \frac{1}{\Gamma(\kappa)} \gamma\left(\kappa, \frac{x}{\theta}\right)$ where $\gamma(\cdot)$ is the lower incomplete gamma function. $F^{-1}(\cdot)$ has no explicit analytical form.	$0 < x < +\infty$ 

Generalized Beta

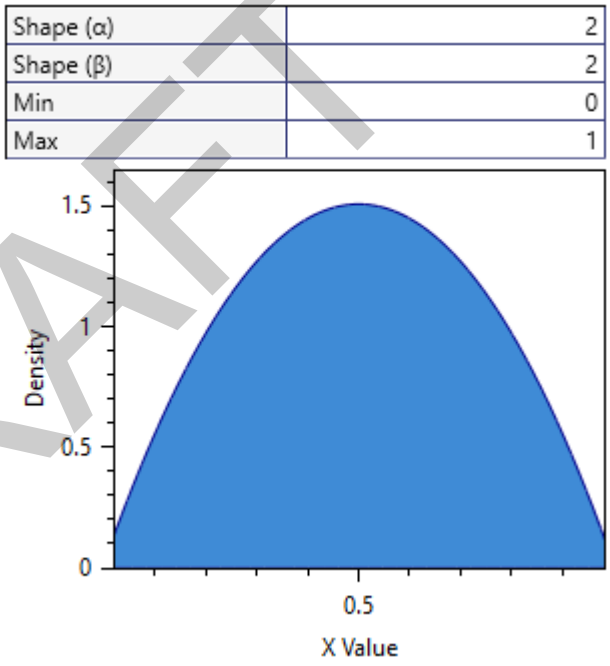
The Generalized Beta distribution, commonly called the four-parameter Beta distribution, is highly flexible and bounded in the closed interval $[a, c]$. The PERT distribution is a transformation of the Generalized Beta. The Generalized Beta distribution provides flexibility to fit expert elicitations or other bounded data.

Table 18 - Summary table for the Generalized Beta distribution.

<table>
<tr><th>Parameters</th><th>Support</th></tr>
<tr><td>Shape: α
Shape: β
Min: a
Max: c</td><td>$\begin{cases} 0 < \alpha < +\infty \\ 0 < \beta < +\infty \\ -\infty < a < c < +\infty \end{cases}$</td></tr>
<tr><th>Distribution Functions</th><th>Domain</th></tr>
<tr><td>$f(x) = \frac{(x-a)^{\alpha-1}(c-x)^{\beta-1}}{B(\alpha,\beta)(c-a)^{\alpha+\beta-1}}$
where $B(\cdot)$ is the beta function.

$F(x) = I(\alpha,\beta,z), \quad z = \frac{x-a}{c-a}$
where $I(\cdot)$ is the incomplete beta function.

$F^{-1}(p) = I^{-1}(\alpha,\beta,p)(c-a)+a$
where $I^{-1}(\cdot)$ is the inverse incomplete beta function.</td><td>$a \le x \le c$

 
</td></tr>
</table>

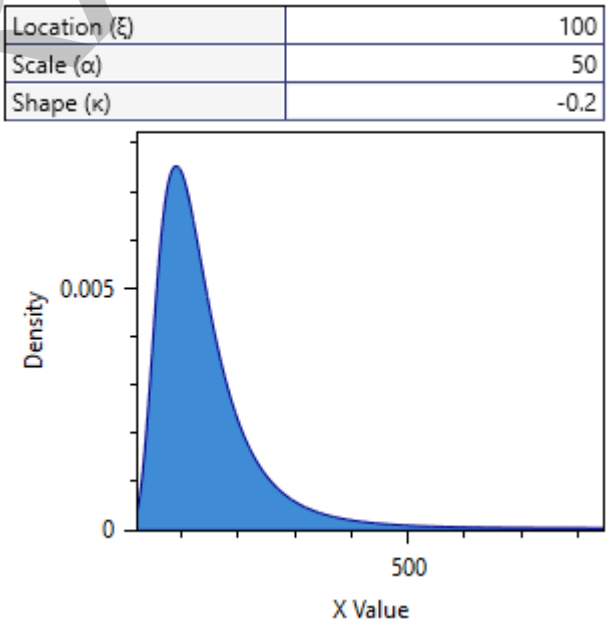
Generalized Extreme Value

The Generalized Extreme Value (GEV) distribution is a three-parameter distribution that subsumes the three extreme-value distributions: Gumbel (EVI), Fréchet (EVII), and Weibull (EVIII). The shape parameter κ determines which sub-distribution the GEV represents. The extreme value theorem states that the GEV distribution is the limit distribution of maxima of a sequence of independent and identically distributed random values. GEV is used for hydrologic frequency analysis, insurance, and financial risks.

RMC software uses Hosking's parameterization [17], in which a GEV with a negative shape parameter κ has no upper bound, and a GEV with a positive κ has an upper bound. Other sources adopt the opposite convention, where a negative shape parameter implies an upper bound.

There are several three parameter distributions available in RMC software. Each have special cases when the shape parameter is equal to zero. To ensure numerical stability, RMC software assumes the shape parameter is sufficiently close to zero when the absolute value of the shape parameter is less than or equal to 10^{-6}. This is consistent with other distribution fitting and frequency analysis software, such as the '*lmom*' and '*nsRFA*' packages in *R* [66] [67].

Table 19 - Summary table for the Generalized Extreme Value distribution.

Parameters	**Support**
Location: ξ Scale: α Shape: κ	$\begin{cases} -\infty < \xi < +\infty \\ 0 < \alpha < +\infty \\ -\infty < \kappa < +\infty \end{cases}$
Distribution Functions	**Domain**
$z = \begin{cases} -\frac{1}{\kappa}\ln\left\{1-\kappa\left(\frac{x-\xi}{\alpha}\right)\right\}, & \kappa \neq 0 \\ \frac{x-\xi}{\alpha}, & \kappa = 0 \end{cases}$ $f(x) = \frac{1}{\alpha}e^{-(1-\kappa)z-e^{-z}}$ $F(x) = e^{-e^{-z}}$ $F^{-1}(p) = \begin{cases} \xi + \frac{\alpha}{\kappa}\{1-(-\ln(p))^{\kappa}\}, & \kappa \neq 0 \\ \xi - \alpha\ln(-\ln(p)), & \kappa = 0 \end{cases}$	$\begin{cases} \xi + \frac{\alpha}{\kappa} \leq x < +\infty, & \kappa < 0 \\ -\infty < x < +\infty, & \kappa = 0 \\ -\infty < x \leq \xi + \frac{\alpha}{\kappa}, & \kappa > 0 \end{cases}$ 

Generalized Logistic

The Generalized Logistic (GLO) distribution is a heavy-tailed, three-parameter distribution. The GLO distribution has been used to fit values of extremes, such as stock return fluctuations and sea levels. It has been used extensively for modeling annual rainfall maxima, and for flood frequency analysis.

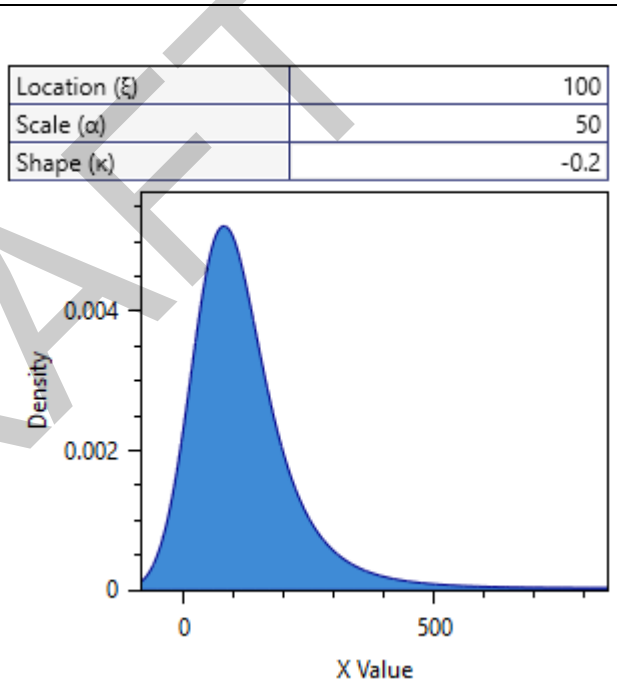
Table 20 - Summary table for the Generalized Logistic distribution.

<table>
<tr><th>Parameters</th><th>Support</th></tr>
<tr><td>Location: ξ
Scale: α
Shape: κ</td><td>$\begin{cases} -\infty < \xi < +\infty \\ 0 < \alpha < +\infty \\ -\infty < \kappa < +\infty \end{cases}$</td></tr>
<tr><th>Distribution Functions</th><th>Domain</th></tr>
<tr><td>$z = \begin{cases} -\frac{1}{\kappa}\ln\left\{1-\kappa\left(\frac{x-\xi}{\alpha}\right)\right\}, & \kappa \neq 0 \\ \frac{x-\xi}{\alpha}, & \kappa = 0 \end{cases}$

$f(x) = \frac{1}{\alpha}e^{-(1-\kappa)z}(1+e^{-z})^{-2}$

$F(x) = (1+e^{-z})^{-1}$

$F^{-1}(p) = \begin{cases} \xi + \frac{\alpha}{\kappa}\left\{1-\left(\frac{1-p}{p}\right)^{\kappa}\right\}, & \kappa \neq 0 \\ \xi - \alpha\ln\left(\frac{1-p}{p}\right), & \kappa = 0 \end{cases}$</td><td>$\begin{cases} \xi + \frac{\alpha}{\kappa} \leq x < +\infty, & \kappa < 0 \\ -\infty < x < +\infty, & \kappa = 0 \\ -\infty < x \leq \xi + \frac{\alpha}{\kappa}, & \kappa > 0 \end{cases}$

 
</td></tr>
</table>

Generalized Normal

The Generalized Normal (GNO) distribution is a flexible three-parameter distribution that has been used extensively for modeling annual rainfall maxima, and for flood frequency analysis. The GNO contains three special cases: 1) a three-parameter log-normal distribution with finite lower bound and positive skewness $(\kappa < 0)$; 2) a two-parameter normal distribution $(\kappa = 0)$; and 3) a reverse three-parameter log-normal distribution with a finite upper bound and negative skewness $(\kappa > 0)$.

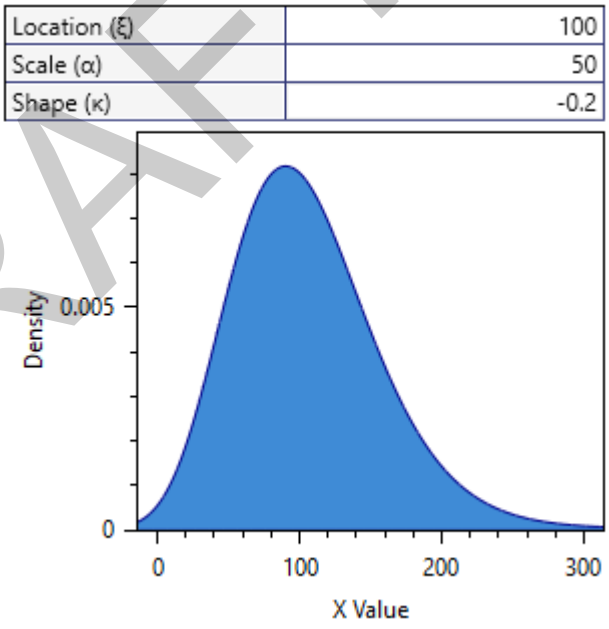
Table 21 - Summary table for the Generalized Normal distribution.

<table>
<tr><th>Parameters</th><th>Support</th></tr>
<tr><td>Location: ξ
Scale: α
Shape: κ</td><td>$\begin{cases} -\infty < \xi < +\infty \\ 0 < \alpha < +\infty \\ -\infty < \kappa < +\infty \end{cases}$</td></tr>
<tr><th>Distribution Functions</th><th>Domain</th></tr>
<tr><td>$z = \begin{cases} -\frac{1}{\kappa}\ln\left\{1-\kappa\left(\frac{x-\xi}{\alpha}\right)\right\}, & \kappa \neq 0 \\ \frac{x-\xi}{\alpha}, & \kappa = 0 \end{cases}$

$f(x) = \frac{e^{\left(\kappa z - \frac{z^2}{2}\right)}}{\alpha\sqrt{2\pi}}$

$F(x) = \Phi(z)$
where $\Phi(\cdot)$ is the Normal CDF.

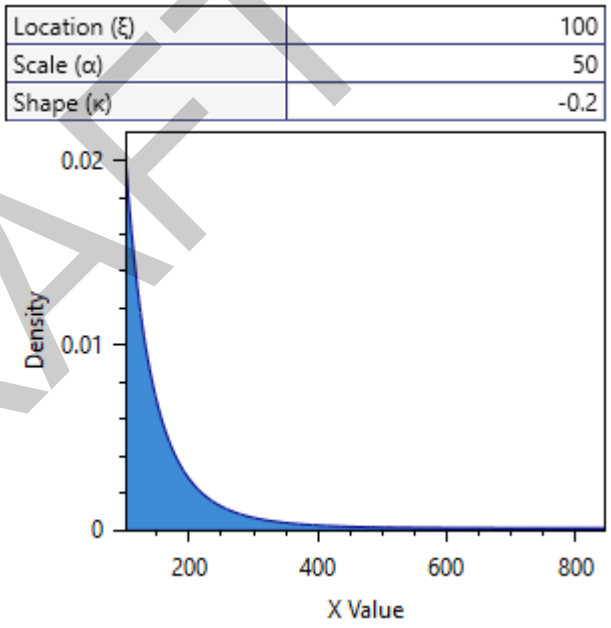
$F^{-1}(p) = \begin{cases} \xi - \frac{\alpha}{\kappa}\left\{e^{-\kappa\cdot\Phi^{-1}(p)} - 1\right\}, & \kappa \neq 0 \\ \xi + \alpha\cdot\Phi^{-1}(p), & \kappa = 0 \end{cases}$
where Φ^{-1} is the Normal inverse CDF.</td><td>$\begin{cases} \xi + \frac{\alpha}{\kappa} \leq x < +\infty, & \kappa < 0 \\ -\infty < x < +\infty, & \kappa = 0 \\ -\infty < x \leq \xi + \frac{\alpha}{\kappa}, & \kappa > 0 \end{cases}$

 
</td></tr>
</table>

Generalized Pareto

The Generalized Pareto (GPA) distribution is a three-parameter distribution with a fixed lower bound. The GPA is upper bounded when the shape parameter is positive. When κ is zero, the distribution reduces to a shifted Exponential distribution. The GPA distribution is most often used with peaks-over-threshold data. Hydrologic frequency analysis uses the GPA in cases where rainfall or flow maxima exceed a specified threshold.

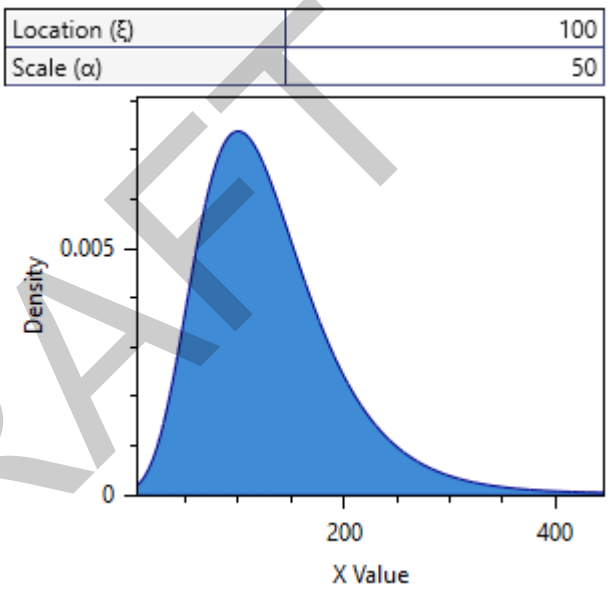
Table 22 - Summary table for the Generalized Pareto distribution.

<table>
<tr><th>Parameters</th><th>Support</th></tr>
<tr><td>Location: ξ
Scale: α
Shape: κ</td><td>$\begin{cases} -\infty < \xi < +\infty \\ 0 < \alpha < +\infty \\ -\infty < \kappa < +\infty \end{cases}$</td></tr>
<tr><th>Distribution Functions</th><th>Domain</th></tr>
<tr><td>$z = \begin{cases} -\frac{1}{\kappa}\ln\left\{1-\kappa\left(\frac{x-\xi}{\alpha}\right)\right\}, & \kappa \neq 0 \\ \frac{x-\xi}{\alpha}, & \kappa = 0 \end{cases}$
$f(x) = \frac{1}{\alpha}e^{-(1-\kappa)z}$
$F(x) = 1 - e^{-z}$
$F^{-1}(p) = \begin{cases} \xi + \frac{\alpha}{\kappa}\{1-(1-p)^{\kappa}\}, & \kappa \neq 0 \\ \xi - \alpha\ln(1-p), & \kappa = 0 \end{cases}$</td><td>$\begin{cases} \xi \leq x < +\infty, & \kappa \leq 0 \\ \xi \leq x \leq \xi + \frac{\alpha}{\kappa}, & \kappa > 0 \end{cases}$
 
</td></tr>
</table>

Gumbel

The Gumbel, or Extreme Value Type-I (EVI) distribution, is a two-parameter distribution with a fixed positive skewness of ≈1.14. The Gumbel distribution is used to describe the maximum (or minimum) of a sample and has seen widespread use in hydrologic frequency analysis. The Gumbel distribution is a particular case of the Generalized Extreme Value (GEV) distribution when the GEV shape parameter is zero. The Gumbel distribution is also used for a probability plotting scale because it exaggerates the extreme tails of the data.

Table 23 - Summary table for the Gumbel distribution.

Parameters	**Support**
Location: ξ Scale: α	$\begin{cases} -\infty < \xi < +\infty \\ 0 < \alpha < +\infty \end{cases}$
Distribution Functions	**Domain**
$z = \frac{x - \xi}{\alpha}$ $f(x) = \frac{1}{\alpha} e^{-(z+e^{-z})}$ $F(x) = e^{-e^{-z}}$ $F^{-1}(p) = \xi - \alpha \ln(-\ln(p))$	$-\infty < x < +\infty$ 

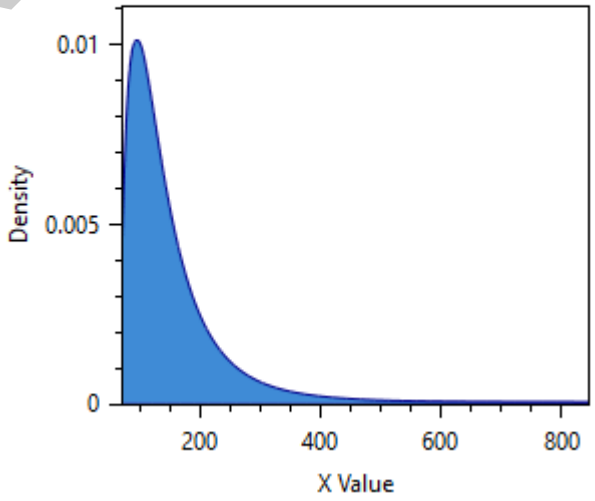
Kappa-4

The Kappa-4 (K4) distribution is a flexible four-parameter distribution that is predominately used for regional rainfall-frequency analysis [17]. The K4 special cases include the Generalized Logistic $(h = -1)$, Generalized Extreme Value $(h = 0)$, Generalized Pareto $(h = 1)$, Logistic $(\kappa = 0, h = -1)$, Gumbel $(\kappa = 0, h = 0)$, Exponential $(\kappa = 0, h = 1)$ and Uniform $(\kappa = 1, h = 1)$ distributions.

Table 24 - Summary table for the Kappa-4 distribution.

Parameters	**Support**
Location: ξ Scale: α Shape: κ Shape: h	$\begin{cases} -\infty < \xi < +\infty \\ 0 < \alpha < +\infty \\ -\infty < \kappa < +\infty \\ -\infty < h < +\infty \end{cases}$
Distribution Functions	**Domain**
$f(x) = \frac{1}{\alpha}\left\{1 - \kappa\left(\frac{x-\xi}{\alpha}\right)\right\}^{\frac{1}{\kappa}-1}\{F(x)\}^{1-h}$ $F(x) = \left[1 - h\left\{1 - \kappa\left(\frac{x-\xi}{\alpha}\right)\right\}^{\frac{1}{\kappa}}\right]^{\frac{1}{h}}$ $F^{-1}(p) = \xi + \frac{\alpha}{\kappa}\left\{1 - \left(\frac{1-p^h}{h}\right)^{\kappa}\right\}$	$\begin{cases} x < +\infty, & \kappa \le 0 \\ x \le \xi + \frac{\alpha}{\kappa}, & \kappa > 0 \\ \xi + \frac{\alpha}{\kappa} \le x, & h \le 0, \kappa < 0 \\ -\infty < x, & h \le 0, \kappa \ge 0 \\ \xi + \frac{\alpha}{\kappa}(1 - h^{-\kappa}) \le x, & h > 0 \end{cases}$

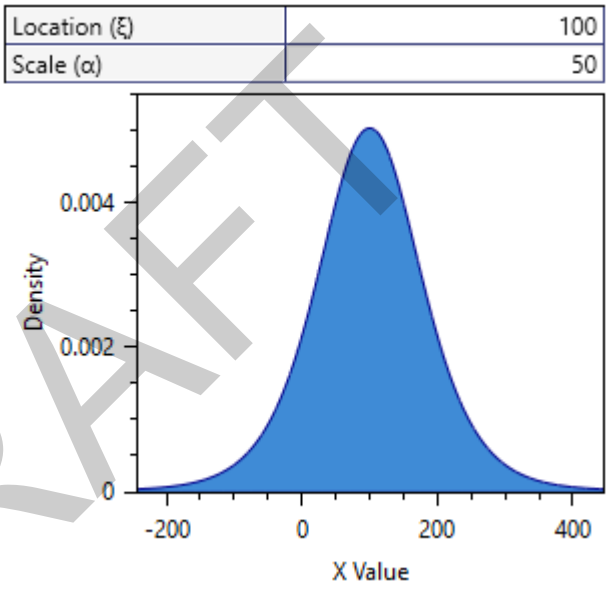
Location (ξ)	100
Scale (α)	50
Shape (κ)	-0.2
Shape (h)	0.5



Logistic

The Logistic distribution is a two-parameter, symmetric distribution with heavier tails (higher kurtosis) than the Normal distribution. The Logistic distribution has applications in hydrology for long duration discharge or rainfall, such as monthly or yearly totals. The most common application is in logistic regression where the errors follow a Logistic distribution.

Table 25 - Summary table for the Logistic distribution.

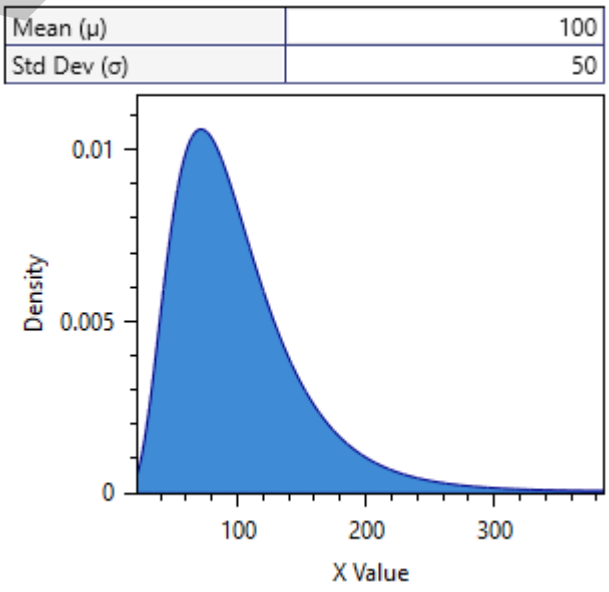
Parameters	**Support**
Location: ξ Scale: α	$\begin{cases} -\infty < \xi < +\infty \\ 0 < \alpha < +\infty \end{cases}$
Distribution Functions	**Domain**
$z = \frac{x - \xi}{\alpha}$ $f(x) = \frac{1}{\alpha} e^{-z}(1 + e^{-z})^{-2}$ $F(x) = (1 + e^{-z})^{-1}$ $F^{-1}(p) = \xi + \alpha \ln\left(\frac{p}{1-p}\right)$	$-\infty < x < +\infty$ 

Log-Normal

The log-Normal distribution is a two-parameter positively skewed distribution that describes a random variable whose logarithm is Normally distributed. The central limit theorem states that a log-Normal process arises from the multiplicative product of many independent random variables, each of which is positive. In hydrology, the log-Normal distribution is used for frequency analysis of annual maximum discharge. In reliability analysis, it is often used to model times to repair a system.

RMC software contains two log-Normal distributions. The first, named “Ln-Normal” is based on the natural logarithm, or log base e. This distribution is parameterized using real-space moments to be more intuitive for multi-disciplinary users of the software. The other distribution, named “Log-Normal” uses log base 10 and is parameterized using $\log_{10}$ moments, which is consistent with typical practice in hydrologic frequency analysis. Both log-Normal distributions are functionally identical and will produce the same statistical inference.

Table 26 - Summary table for the Log-Normal distribution.

<table>
<tr><th>Parameters</th><th>Support</th></tr>
<tr><td>Location: μ
Scale: σ
Conversion from real-space to Ln-space:
$\mu = \ln\left(\frac{\mu_0^2}{\sqrt{\mu_0^2 + \sigma_0^2}}\right)$
$\sigma = \sqrt{\ln\left(1 + \frac{\sigma_0^2}{\mu_0^2}\right)}$</td><td>$\begin{cases} -\infty < \mu < +\infty \\ 0 < \sigma < +\infty \end{cases}$</td></tr>
<tr><th>Distribution Functions</th><th>Domain</th></tr>
<tr><td>$z = \frac{\ln x - \mu}{\sigma}$
$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$
$F(x) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{1}{\sqrt{2}}z\right)\right]$
where erf is the error function.
$F^{-1}(p) = e^{\mu + \sigma\sqrt{2}\,\mathrm{erf}^{-1}(2p-1)}$
where erf^{-1} is the inverse error function.</td><td>$-\infty < x < +\infty$
 
</td></tr>
</table>

Log-Pearson Type III

The log-Pearson Type III (LPIII) distribution is a flexible three-parameter distribution that describes a random variable whose logarithm is PIII distributed. The LPIII distribution was originally used to model annual maximum flood flows in the United States [68].

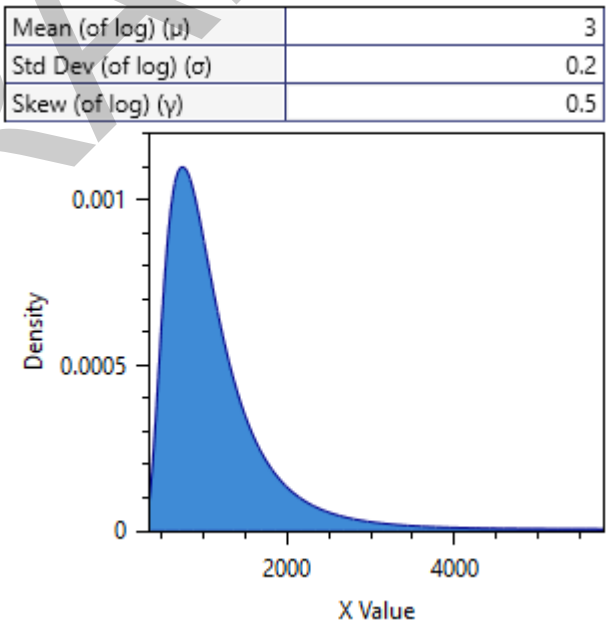
In RMC software, the LPIII uses log base 10 and is parameterized using $\log_{10}$ central moments of the distribution mean (μ), standard deviation (σ), and skewness (γ). The true parameters (location, scale, and shape) are computed from the specified moments. This is done because the moments of the distribution are more intuitively defined by end-users familiar with Bulletin 17B [69] and Bulletin 17C [70]. RMC software uses the same parameterization as [17], with the underlying location parameter ξ, the scale parameter β, and the shape parameter α.

Table 27 - Summary table for the Log-Pearson Type III distribution.

<table>
<tr><th>Parameters</th><th>Support</th></tr>
<tr><td>Location: $\xi = \mu - \frac{2\sigma}{\gamma}$
Scale: $\beta = \frac{1}{2}\sigma\gamma$
Shape: $\alpha = \frac{4}{\gamma^2}$</td><td>$\begin{cases} -\infty < \mu < +\infty \\ 0 < \sigma < +\infty \\ -\infty < \gamma < +\infty \end{cases}$
The method of maximum likelihood requires that $-2 \le \gamma \le +2$</td></tr>
<tr><th>Distribution Functions</th><th>Domain</th></tr>
<tr><td>$f(x) = \begin{cases} g(\log_{10}(\xi - x)\,\vert\beta,\alpha)\left(\frac{K}{x}\right), & \gamma < 0 \\ \phi(\log_{10}(x)\,\vert\mu,\sigma)\left(\frac{K}{x}\right), & \gamma = 0 \\ g(\log_{10}(x - \xi)\,\vert\beta,\alpha)\left(\frac{K}{x}\right), & \gamma > 0 \end{cases}$
where $g(\cdot)$ is the Gamma PDF; and $\phi(\cdot)$ is the Normal PDF; and $K = \log_{10} e$.

$F(x) = \begin{cases} 1 - G(\log_{10}(\xi - x)\,\vert\beta,\alpha), & \gamma < 0 \\ \Phi(\log_{10}(x)\,\vert\mu,\sigma), & \gamma = 0 \\ G(\log_{10}(x - \xi)\,\vert\beta,\alpha), & \gamma > 0 \end{cases}$
where $G(\cdot)$ is the Gamma CDF; and $\Phi(\cdot)$ is the Normal CDF.

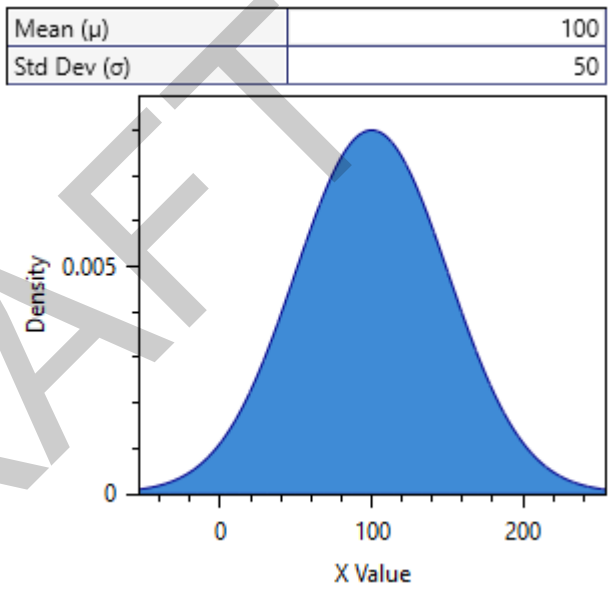
$F^{-1}(p) = \begin{cases} e^{\left(\xi - G^{-1}(1-p\vert\beta,\alpha)\right)/K}, & \gamma < 0 \\ e^{\Phi^{-1}(p\vert\mu,\sigma)/K}, & \gamma = 0 \\ e^{\left(\xi + G^{-1}(p\vert\beta,\alpha)\right)/K}, & \gamma > 0 \end{cases}$
where G^{-1} is the Gamma inverse CDF; and Φ^{-1} is the Normal inverse CDF.</td><td>$\begin{cases} 0 < x \le \xi, & \gamma < 0 \\ 0 < x < +\infty, & \gamma = 0 \\ \xi \le x < +\infty, & \gamma > 0 \end{cases}$

 
</td></tr>
</table>

Normal

The Normal (or Gaussian) distribution is a very widely used two-parameter probability distribution. The Normal distribution is fundamental to most statistical modeling because of the central limit theorem, which states that the mean of independent random variables trends towards a Normal distribution, even if the original variables themselves are not Normally distributed. The Normal distribution fits many natural phenomena, such as body heights, blood pressure, measurement error, and annual rainfall.

Table 28 - Summary table for the Normal distribution.

Parameters	**Support**
Location: μ Scale: σ	$\begin{cases} -\infty < \mu < +\infty \\ 0 < \sigma < +\infty \end{cases}$
Distribution Functions	**Domain**
$z = \frac{x-\mu}{\sigma}$ $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$ $F(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{1}{\sqrt{2}}z\right)\right]$ where erf is the error function. $F^{-1}(p) = \mu + \sigma\sqrt{2}\operatorname{erf}^{-1}(2p-1)$ where erf^{-1} is the inverse error function.	$-\infty < x < +\infty$ 

Nonparametric

In many cases, probability distributions for the hazard and system response will be nonparametric, either based on empirical data, simulation results, or expert elicitation. The nonparametric distributions functions are based on linear interpolation as shown below. There is often a need to transform the x values or probability values to improve the accuracy of the linear interpolation. For example, the log-linear interpolation is the same as linear interpolation, only it takes place over the logarithm of the x values.

Table 29 - Summary table for the Nonparametric distribution.

Parameters	**Support**
Values $\{x\} = \{x_1, x_2, \dots, x_n\}$ Probabilities $\{p\} = \{p_1, p_2, \dots, p_n\}$	$\begin{cases} -\infty < x_i < +\infty \\ 0 \le p_i \le 1 \end{cases}$
Distribution Functions	**Domain**
$f(x) = \frac{p_{i+1} - p_i}{x_{i+1} - x_i}$ $F(x) = p_i + (p_{i+1} - p_i)\left(\frac{x - x_i}{x_{i+1} - x_i}\right)$ $F^{-1}(p) = x_i + (x_{i+1} - x_i)\left(\frac{p - p_i}{p_{i+1} - p_i}\right)$	$x_1 \le x \le x_n$ The $\{x\}$ values and non-exceedance probabilities $\{p\}$ must be sorted in ascending order $x_i < x_{i+1}$ and $p_i \le p_{i+1}$.

Pearson Type III

The Pearson Type III (PIII) distribution is a three-parameter distribution that is widely used in hydrologic frequency analysis. It has also been used to model the probability of wind speed and rainfall intensity. The PIII distribution is deduced from the two-parameter gamma distribution and converges to a Normal distribution as its skewness (γ) approaches zero.

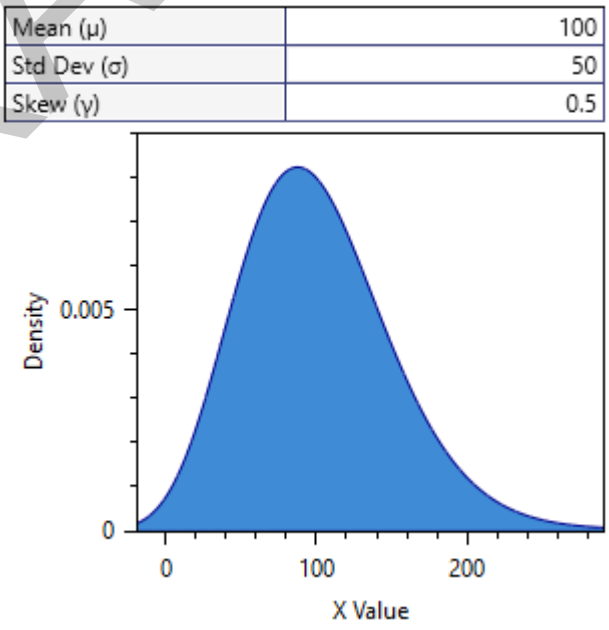
In RMC software, the PIII distribution is parameterized using the central moments of the distribution mean (μ), standard deviation (σ), and skewness (γ). The true parameters (location, scale, and shape) are computed from the specified moments. This is done because the moments of the distribution are more intuitively defined by end-users familiar with Bulletin 17B [69] and Bulletin 17C [70]. RMC software uses the same parameterization as [17], with the underlying location parameter ξ, the scale parameter β, and the shape parameter α.

Table 30 - Summary table for the Pearson Type III distribution.

<table>
<tr><th>Parameters</th><th>Support</th></tr>
<tr><td>Location: $\xi = \mu - \frac{2\sigma}{\gamma}$
Scale: $\beta = \frac{1}{2}\sigma\gamma$
Shape: $\alpha = \frac{4}{\gamma^2}$</td><td>$\begin{cases} -\infty < \mu < +\infty \\ 0 < \sigma < +\infty \\ -\infty < \gamma < +\infty \end{cases}$
The method of maximum likelihood requires that $-2 \leq \gamma \leq +2$</td></tr>
<tr><th>Distribution Functions</th><th>Domain</th></tr>
<tr><td>$f(x) = \begin{cases} g(\xi - x\|\beta,\alpha), & \gamma < 0 \\ \phi(x\|\mu,\sigma), & \gamma = 0 \\ g(x - \xi\|\beta,\alpha), & \gamma > 0 \end{cases}$
where $g(\cdot)$ is the Gamma PDF; and $\phi(\cdot)$ is the Normal PDF.

$F(x) = \begin{cases} 1 - G(\xi - x\|\beta,\alpha), & \gamma < 0 \\ \Phi(x\|\mu,\sigma), & \gamma = 0 \\ G(x - \xi\|\beta,\alpha), & \gamma > 0 \end{cases}$
where $G(\cdot)$ is the Gamma CDF; and $\Phi(\cdot)$ is the Normal CDF.

$F^{-1}(p) = \begin{cases} \xi - G^{-1}(1 - p\|\beta,\alpha), & \gamma < 0 \\ \Phi^{-1}(p\|\mu,\sigma), & \gamma = 0 \\ \xi + G^{-1}(p\|\beta,\alpha), & \gamma > 0 \end{cases}$
where G^{-1} is the Gamma inverse CDF; and Φ^{-1} is the Normal inverse CDF.</td><td>$\begin{cases} 0 < x \leq \xi, & \gamma < 0 \\ 0 < x < +\infty, & \gamma = 0 \\ \xi \leq x < +\infty, & \gamma > 0 \end{cases}$

 
</td></tr>
</table>

PERT

The PERT distribution is a continuous distribution with a closed interval $[a, c]$ and with a mode of b. It is commonly used with expert elicitation when the range is known and there is a reasonable estimate for the most likely value. The PERT distribution is widely used in project management and event tree analysis.

This distribution was first used for estimating project task durations using the *program evaluation and review technique*, which is where the distribution gets its name. The PERT is a transformation of the Generalized Beta (four-parameter Beta) distribution.

The PERT can also be defined based on user-defined 5th, 50th and 95th percentiles (in real-space or Normal z-space). The underlying parameters are solved to minimize the sum of squared errors between the estimated and user-defined percentiles. It should be noted that there might be deviations between the estimated and user-defined percentiles due to convergence limitations of the underlying numerical optimization method.

Table 31 - Summary table for the PERT distribution.

<table>
<tr><th>Parameters</th><th>Support</th></tr>
<tr><td>Min: a
Mode: b
Max: c</td><td>$-\infty < a < b < c < +\infty$</td></tr>
<tr><th>Distribution Functions</th><th>Domain</th></tr>
<tr><td>$\mu = \frac{a + 4b + c}{6}$

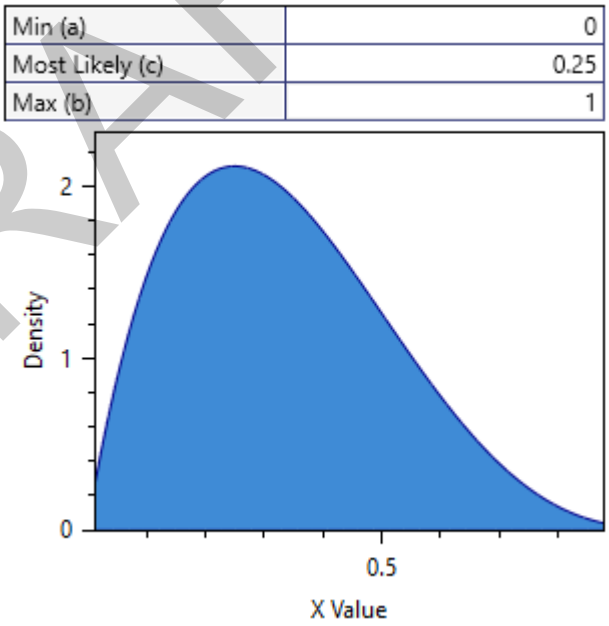
$\alpha = \frac{(\mu - a)(2b - a - c)}{(b - \mu)(c - a)}$

$\beta = \frac{\alpha(c - \mu)}{(\mu - a)}$

$f(x) = \frac{(x - a)^{\alpha-1}(c - x)^{\beta-1}}{B(\alpha, \beta)(c - a)^{\alpha+\beta-1}}$
where $B(\cdot)$ is the beta function.

$F(x) = I(\alpha, \beta, z), \quad z = \frac{x-a}{c-a}$
where $I(\cdot)$ is the incomplete beta function.

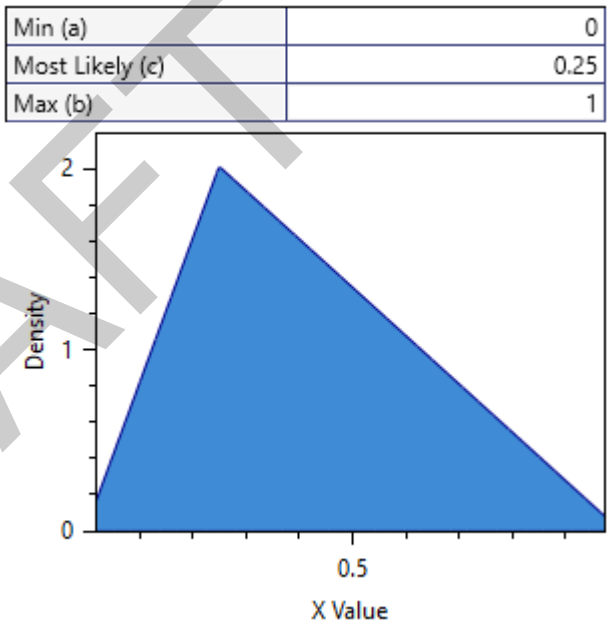
$F^{-1}(p) = I^{-1}(\alpha, \beta, p)(c - a) + a$
where $I^{-1}(\cdot)$ is the inverse incomplete beta function.</td><td>$a \leq x \leq c$

 
</td></tr>
</table>

Triangular

The triangular distribution is a continuous distribution with a closed interval $[a, c]$ and with a mode of b. The triangular distribution is commonly used with expert elicitation when the range is known and there is a reasonable estimate for the most likely value. The triangular distribution, along with the PERT distribution, is widely used in project management and event tree analysis.

Table 32 - Summary table for the Triangular distribution.

<table>
<tr><th>Parameters</th><th>Support</th></tr>
<tr><td>Min: a
Mode: b
Max: c</td><td>$-\infty < a \leq b \leq c \ < +\infty$</td></tr>
<tr><th>Distribution Functions</th><th>Domain</th></tr>
<tr><td>$$f(x) = \begin{cases} \frac{2(x-a)}{(c-a)(b-a)}, & a \leq x \leq b \\ \frac{2(c-x)}{(c-a)(c-b)}, & b < x \leq c \end{cases}$$
$$F(x) = \begin{cases} \frac{(x-a)^2}{(c-a)(b-a)}, & a < x \leq b \\ 1 - \frac{(c-x)^2}{(c-a)(c-b)}, & b < x < c \end{cases}$$
$$F^{-1}(p) = \begin{cases} a + \sqrt{p(c-a)(b-a)}, & p < \frac{b-a}{c-a} \\ c - \sqrt{(1-p)(c-a)(c-b)}, & p \geq \frac{b-a}{c-a} \end{cases}$$</td><td>$a \leq x \leq c$
 
</td></tr>
</table>

Truncated Normal

The Truncated Normal distribution is a Normal distribution bounded by the closed interval $[a, b]$. The Truncated Normal distribution has wide applications in statistics, econometrics, and risk analysis. This distribution provides flexibility to fit expert elicitations like the Generalized Beta distribution.

Table 33 - Summary table for the Truncated Normal distribution.

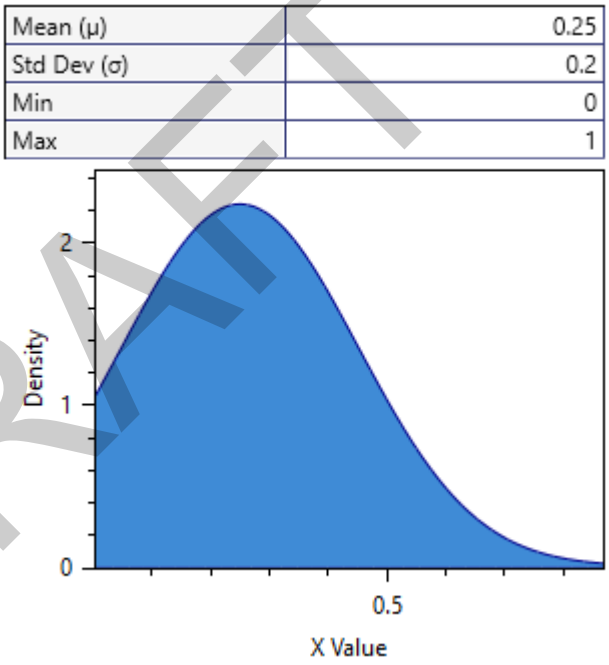
<table>
<tr><th>Parameters</th><th>Support</th></tr>
<tr><td>Location: μ
Scale: σ
Min: a
Max: b</td><td>$\begin{cases} -\infty < \mu < +\infty \\ 0 < \sigma < +\infty \\ -\infty < a < b < +\infty \end{cases}$</td></tr>
<tr><th>Distribution Functions</th><th>Domain</th></tr>
<tr><td>$\xi = \frac{x-\mu}{\sigma}, \quad \alpha = \frac{a-\mu}{\sigma}, \quad \beta = \frac{b-\mu}{\sigma}$

$Z = \Phi(\alpha) - \Phi(\beta)$

$f(x) = \frac{\phi(\xi)}{\sigma Z}$
where $\phi(\cdot)$ is the Normal PDF; and $\Phi(\cdot)$ is the Normal CDF.

$F(x) = \frac{\Phi(\xi)-\Phi(\alpha)}{Z}$

$F^{-1}(p) = \mu + \sigma\Phi^{-1}(\Phi(\alpha) + pZ)$
where Φ^{-1} is the Normal inverse CDF.</td><td>$a \leq x \leq b$

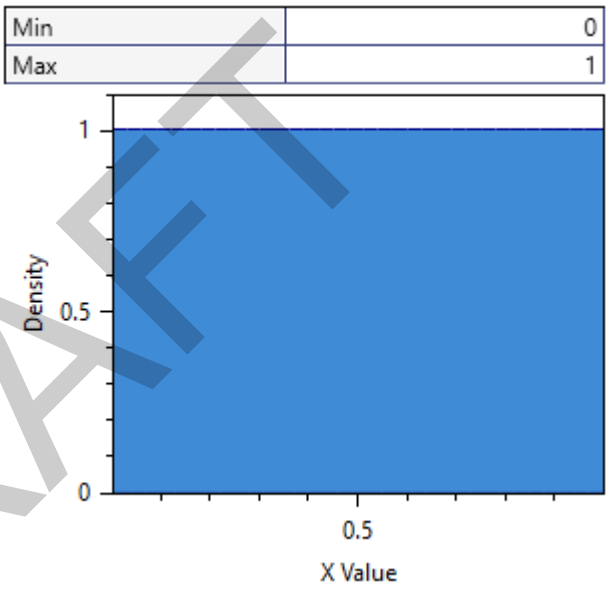
 
</td></tr>
</table>

Uniform

The uniform distribution is a symmetrical distribution that assigns equal probability density to all outcomes in the closed interval $[a, b]$. The uniform distribution is widely used for generating random numbers for a Monte Carlo simulation. It is also used in Bayesian analysis for a default, uninformative prior distribution. It can be used for expert elicitation when no information about the mode exists, but the range is known.

Table 34 - Summary table for the Uniform distribution.

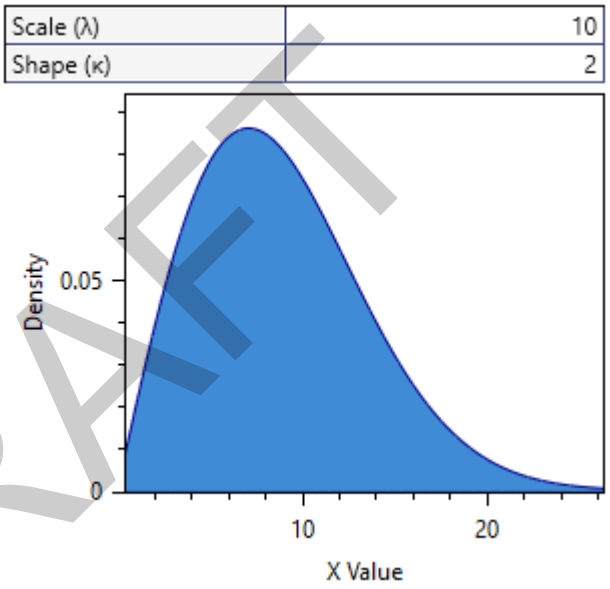
Parameters	**Support**
Min: a Max: b	$-\infty < a < b < +\infty$
Distribution Functions	**Domain**
$f(x) = \frac{1}{b-a}$ $F(x) = \frac{x-a}{b-a}$ $F^{-1}(p) = a + p(b-a)$	$a \leq x \leq b$



Weibull

The Weibull distribution is a two-parameter distribution commonly used in reliability analysis. It is related to several other probability distributions. In particular, the Weibull distribution interpolates between the exponential distribution (for $\kappa = 1$) and the Rayleigh distribution (when $\kappa = 2$). If the quantity x is the "time to failure", the Weibull distribution gives the distribution for which the failure rate is proportional to a power of time.

Table 35 - Summary table for the Weibull distribution.

Parameters	**Support**
Scale: λ Shape: κ	$\begin{cases} 0 < \lambda < +\infty \\ 0 < \kappa < +\infty \end{cases}$
Distribution Functions	**Domain**
$f(x) = \frac{\kappa}{\lambda}\left(\frac{x}{\lambda}\right)^{\kappa-1} e^{-\frac{x^\kappa}{\lambda}}$ $F(x) = 1 - e^{-\frac{x^\kappa}{\lambda}}$ $F^{-1}(p) = \lambda \ln\left(\frac{1}{1-p}\right)^{\frac{1}{\kappa}}$	$0 \leq x < +\infty$ 

Appendix F – Multivariate Normal Distribution

In RMC-TotalRisk, the dependency between failure modes and system components is modeled with the Multivariate Normal (MVN) distribution, which is often referred to as a *Gaussian Copula*. The following details on the MVN are taken from [16] and [71]. Whether a random vector $\boldsymbol{x} = (x_1, \cdots, x_k)^T$ is normally distributed or not, it has a mean vector:

$$\boldsymbol{\mu} = \mathbb{E}[\boldsymbol{x}] = (\mathbb{E}[x_1], \cdots, \mathbb{E}[x_k])^T \quad \textit{Equation 199}$$

and a $k \times k$ covariance matrix:

$$\boldsymbol{\Sigma} = \mathbb{E}[(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^T] \quad \textit{Equation 200}$$

where $(\cdot)^T$ represents the transposed vector, or column vector. The off-diagonal elements in $\boldsymbol{\Sigma}$ are the covariances, and the diagonal elements are the variances.

$$\boldsymbol{\Sigma}_{ij} = \mathrm{cov}(x_i, x_j) = \sigma_{ij} \quad \textit{Equation 201}$$

$$\boldsymbol{\Sigma}_{ii} = \mathrm{var}(x_i) = \sigma_{ii} \quad \textit{Equation 202}$$

The off-diagonal elements in $\boldsymbol{\Sigma}$ relate to the correlations between the coordinates of $\boldsymbol{x}$:

$$\mathrm{cor}(x_i, x_j) = \rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}} \quad \textit{Equation 203}$$

The correlation coefficient ρ between two variables must have a value between $+1$ and -1. When $k > 2$, there are further constraints on the magnitude of ρ (see the **Perfect Negative Dependency** section below).

The probability density function of the MVN is:

$$f(x_1, \cdots, x_k) = \frac{e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{u})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{u})}}{\sqrt{(2\pi)^k |\boldsymbol{\Sigma}|}} \quad \textit{Equation 204}$$

where $\boldsymbol{x}$ is a k-dimensional vector of values; $\boldsymbol{u}$ is a k-dimensional vector of mean values; $\boldsymbol{\Sigma}$ is the $k \times k$ covariance matrix; and $|\boldsymbol{\Sigma}|$ is the determinant of $\boldsymbol{\Sigma}$, also know as the generalized variance.

It is important to note that while the dependency between failure modes and system components is modeled with the MVN, their marginal distributions are not required to be Normally distributed. The probability values of each marginal distribution are transformed to standard Normal z variates. Then, the dependency between the transformed z variates is modeled with the MVN, assuming a mean vector of 0 and variances of 1. When the variances are 1, the covariance matrix and correlation matrix are equivalent. Please see the Correlation Matrix section below for more details.

Positive Definite Matrices

The covariance matrix $\boldsymbol{\Sigma}$ of the MVN must be positive definite. For a symmetric matrix $\boldsymbol{A}$ to be positive definite, the determinant must be greater than zero. The determinant describes important information about the matrix with just a single scalar value. For example, the determinant of a 2×2 matrix is:

$$|\boldsymbol{A}| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc \qquad \textit{Equation 205}$$

And the determinant of a 3×3 matrix is:

$$|\boldsymbol{A}| = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix} \qquad \textit{Equation 206}$$

$$|\boldsymbol{A}| = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei + bfg + cdh - ceg - bdi - afh \qquad \textit{Equation 207}$$

There are a few important qualities of a positive definite matrix [71]:

- If matrix $\boldsymbol{A}$ is positive definite, then its eigenvalues are positive numbers and therefore the matrix is invertible.
- The inverse of a positive definite matrix is also a positive definite matrix.
- If matrix $\boldsymbol{A}$ is positive definite, then all the diagonal entries are positive.
- If matrix $\boldsymbol{A}$ is positive definite, then there exists a unique lower triangular matrix $\boldsymbol{R}$, with positive diagonal entries, such that:

$$\boldsymbol{A} = \boldsymbol{R}\boldsymbol{R}^T \qquad \textit{Equation 208}$$

Equation 208 plays an important role in Cholesky decomposition, which is required to generate random values from a MVN distribution.

Correlation Matrix

The covariance matrix for three variables can be written as:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} & \sigma_{1,3} \\ \sigma_{2,1} & \sigma_{2,2} & \sigma_{2,3} \\ \sigma_{3,1} & \sigma_{3,2} & \sigma_{3,3} \end{bmatrix} \qquad \textit{Equation 209}$$

where the off-diagonal elements are the covariances, $\sigma_{1,2}$, and the diagonal elements are the variances, $\sigma_{1,1} = \sigma_1^2$. When using the standard MVN for dependency, the mean vector is 0 and variances are 1. As such, the covariance and correlation between two variables are equivalent:

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}} = \frac{\sigma_{ij}}{\sqrt{1 \cdot 1}} = \sigma_{ij} \qquad \textit{Equation 210}$$

For this reason, the dependency structure in RMC-TotalRisk is defined using a correlation matrix rather than a covariance matrix.

Because there is typically limited observed data on failure modes, there are only two options for defining dependency between failure modes within a system component. As discussed in the Quantitative Risk Analysis chapter, by default every failure mode is assumed to be independent of all the others. This is the unimodal upper bound for a series system when failure modes have non-negative dependency [9]. Alternatively, the user can set each failure mode as non-independent. In this case, these non-independent failure modes are treated as perfectly negatively dependent of all the others. Perfect negative dependency is the unimodal upper bound for series systems [9].

There are four options for defining dependency between system components: 1) perfect independence; 2) perfect positive dependency; 3) perfect negative dependency; and 4) user-defined correlation matrix.

The only requirement for the user-defined correlation matrix is that it be positive definite. The details of the first three options are described in the following subsections.

Perfect Independence

When inputs are perfectly independent, the correlation between inputs is 0. An example correlation matrix for three variables with perfect independence is:

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \textit{Equation 211}$$

Recall from Appendix A, the joint probability of three independent variables is simply:

$$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C) \qquad \textit{Equation 212}$$

For example, if $P(A) = 0.25$, $P(B) = 0.35$, and $P(C) = 0.50$, then:

$$P(A \cap B \cap C) = 0.25 \cdot 0.35 \cdot 0.50 = 0.04375 \qquad \textit{Equation 213}$$

Perfect Positive Dependency

When inputs are perfectly positively dependent, the correlation between inputs is 1. An example correlation matrix for three variables with perfect positive dependency is:

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \qquad \textit{Equation 214}$$

The joint probability of three perfectly positively dependent variables is simply the minimum of each:

$$P(A \cap B \cap C) = \min[P(A), P(B), P(C)] \qquad \textit{Equation 215}$$

In keeping with the previous example:

$$P(A \cap B \cap C) = \min[0.25, 0.35, 0.50] = 0.25 \qquad \textit{Equation 216}$$

Perfect Negative Dependency

When two inputs are perfectly negatively dependent, the correlation between inputs is -1. However, when there are more than two variables, there is a limit to how negative the dependency can be such that the correlation matrix is still positive definite. For example, the correlation matrix for three variables can be written as:

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & a & b \\ a & 1 & c \\ b & c & 1 \end{bmatrix} \quad \textit{Equation 217}$$

The determinant of this matrix must be positive:

$$|\boldsymbol{\Sigma}| = \begin{vmatrix} 1 & a & b \\ a & 1 & c \\ b & c & 1 \end{vmatrix} = 1 + 2abc - a^2 - b^2 - c^2 > 0 \quad \textit{Equation 218}$$

If each variable has the same level of negative correlation, the minimum allowable correlation coefficient ρ can be determined as follows:

$$|\boldsymbol{\Sigma}| = \begin{vmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{vmatrix} = 1 + 2\rho^3 - 3\rho^2 = 0 \quad \textit{Equation 219}$$

Rearranging:

$$2\rho^3 - 3\rho^2 + 1 = 0 \quad \textit{Equation 220}$$

Factoring:

$$(\rho - 1)^2(2\rho + 1) = 0 \quad \textit{Equation 221}$$

Following the zero-factor principle:

$$\rho - 1 = 0 \text{ or } 2\rho + 1 = 0 \quad \textit{Equation 222}$$

Solving each side gives:

$$\rho - 1 = 0 \text{ so } \rho = 1 \quad \textit{Equation 223}$$

$$2\rho + 1 = 0 \text{ so } \rho = -\tfrac{1}{2} \quad \textit{Equation 224}$$

This can also be seen by plotting the determinant function as shown in Figure 75 below. In summary, when each variable has the same level of negative correlation, the formula for the minimum allowable negative correlation can be generalized to be:

$$\rho > -\frac{1}{k-1} \quad \textit{Equation 225}$$

where k is the number of variables or dimensions.

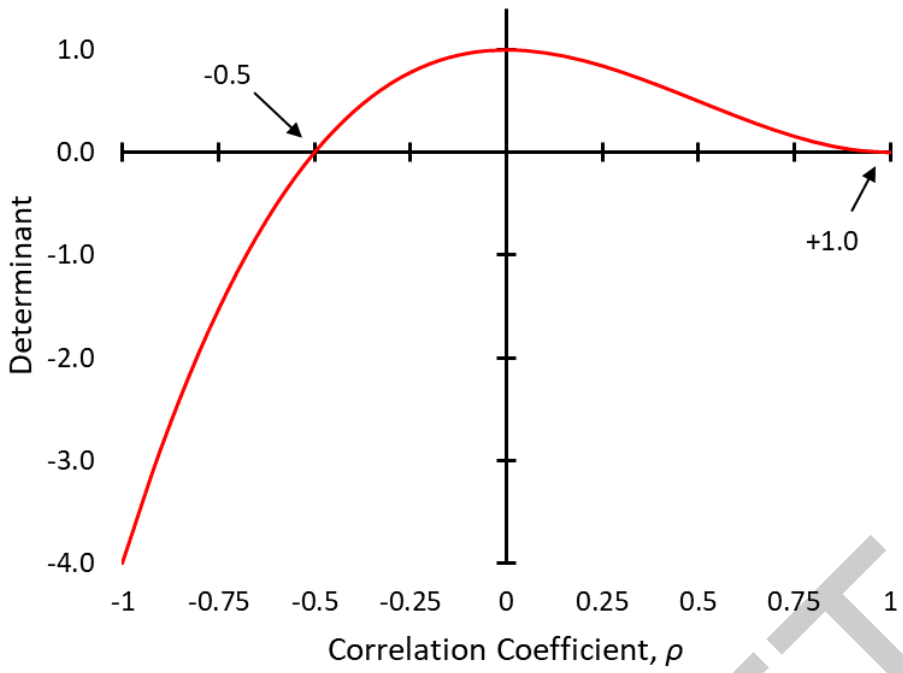


Figure 75 – Graphical function of the determinant of a three variable correlation matrix.

In RMC-TotalRisk, by default every failure mode is assumed to be independent. However, the user can choose to set a single failure mode to be non-independent. In this case, the single non-independent failure mode is set to have perfect negative dependency, while the other failure modes are still assumed to be independent. In keeping with the three-variable example, the user might select variable C to be non-independent. The minimum allowable negative correlation can be solved as follows:

$$|\boldsymbol{\Sigma}| = \begin{vmatrix} 1 & 0 & \rho \\ 0 & 1 & \rho \\ \rho & \rho & 1 \end{vmatrix} = 1 - 2\rho^2 = 0 \qquad \textit{Equation 226}$$

$$|\boldsymbol{\Sigma}| = \begin{vmatrix} 1 & 0 & \rho \\ 0 & 1 & \rho \\ \rho & \rho & 1 \end{vmatrix} = 1 - (k-1)\rho^2 = 0 \qquad \textit{Equation 227}$$

$$\rho = \pm \frac{1}{\sqrt{k-1}} \qquad \textit{Equation 228}$$

A plot of the determinant function in Equation 227 is shown in Figure 76. From this solution, it can be deduced that when only one variable is negatively dependent, while the remaining are independent, the minimum allowable negative correlation is:

$$\rho > -\frac{1}{\sqrt{k-1}} \qquad \textit{Equation 229}$$

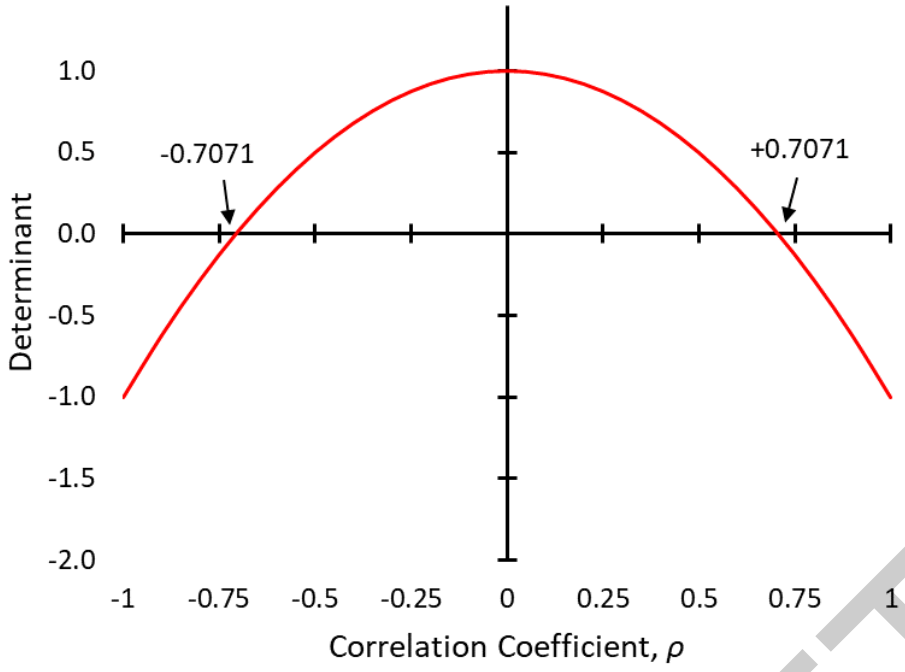


Figure 76 – Graphical function of the determinant of a three variable correlation matrix with one non-independent variable.

Now, consider a scenario with four variables, two are negatively dependent and two are independent. To solve for the minimum allowable negative correlation, set the determinant equal to zero and solve the root.

$$|\boldsymbol{\Sigma}| = \begin{vmatrix} 1 & 0 & \rho & \rho \\ 0 & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{vmatrix} = 0 \qquad \textit{Equation 230}$$

$$-\frac{1}{\sqrt{k-1}} \leq \rho \leq -\frac{1}{k-1} \qquad \textit{Equation 231}$$

$$-0.577 \leq \rho \leq -0.333 \ldots \qquad \textit{Equation 232}$$

After solving the root using the Brent method [64], it is determined that the minimum allowable correlation $\rho > -0.39$.

Computing the joint probability of perfectly negatively dependent variables is challenging. The next section provides details on computing the cumulative distribution function (i.e., the joint probability) of the MVN.

Joint Probability

The risk analysis of a system with several potential failure modes requires the integration of the MVN distribution as follows:

$$F(c_1,\cdots,c_k)=P\left[\bigcap_{i=1}^{k}(X_i\le c_i)\right]=\int_{-\infty}^{c_1}\cdots\int_{-\infty}^{c_k}\frac{e^{-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x}}}{\sqrt{(2\pi)^k|\boldsymbol{\Sigma}|}}\cdot dx_1\cdots dx_k \qquad \textit{Equation 233}$$

where $X_k\le c_k$ denotes k limit states of a system in standard Normal z space, and $\boldsymbol{\Sigma}$ is the correlation matrix among failure modes.

Genz [72] solved Equation 233 with an adaptive Monte Carlo integration technique, and his solution is widely used in statistical software. However, numerical integration becomes impractical for large k due to long runtimes. *Pandey* [73] proposed a simpler approach using the product of conditional marginals (PCM), which has much faster runtimes and maintains a high level of accuracy. *Yuan* and *Pandey* [74] proposed improvements to PCM and demonstrated the superior accuracy of PCM compared to other approximations. Considering the faster runtimes and accuracy, RMC-TotalRisk uses the PCM method for computing the joint probabilities with the MVN distribution.

Product of Conditional Marginals

The main concept behind PCM is that Equation 233 can be represented as a product of k conditional probability terms, where each conditional probability is approximated by a one-dimensional Normal distribution:

$$F(c_1,\cdots,c_k)=P\left[(X_k\le c_k)|\bigcap_{i=1}^{k-1}(X_i\le c_i)\right]\times P\left[(X_{k-1}\le c_{k-1})|\bigcap_{i=1}^{k-2}(X_i\le c_i)\right]\times\cdots\times P(X_1\le c_1) \qquad \textit{Equation 234}$$

$$F(c_1,\cdots,c_k)\approx\Phi\left(c_{k|k-1}\right) \qquad \textit{Equation 235}$$

where $c_{k|k-1}$ represents a conditional Normal fractile (z variate) obtained by assuming that the conditional variable $X_{k|k-1}$ is Normally distributed with a certain mean and variance. Equation 235 is solved in the following way:

$$c_{j|k}=\frac{\left(c_j-\mu_{j|k}\right)}{\sigma_{j|k}} \qquad \textit{Equation 236}$$

where the mean and standard deviation are calculated by the following:

$$\mu_{j|k}=-\rho_{(k,j)|(k-1)}\cdot A_{k|(k-1)} \qquad \textit{Equation 237}$$

$$\sigma_{j|k}=\sqrt{1-\rho^2{}_{(k,j)|(k-1)}\cdot B_{k|(k-1)}} \qquad \textit{Equation 238}$$

$$A_{k|(k-1)}=\frac{\phi\left(c_{k|k-1}\right)}{\Phi\left(c_{k|k-1}\right)} \qquad \textit{Equation 239}$$

$$B_{k|(k-1)}=A_{k|(k-1)}\cdot\left(c_{k|k-1}+A_{k|(k-1)}\right) \qquad \textit{Equation 240}$$

The correlation between X_k and X_j conditional on $\bigcap_{i=1}^{k-1}(X_i \le c_i)$ is given by:

$$\rho_{(k+1,j+1)|k} = \frac{\rho_{(k+1,j+1)|(k-1)} - \rho_{(k,k+1)|(k-1)} \cdot \rho_{(k,j+1)|(k-1)} \cdot B_{k|(k-1)}}{\sqrt{\left[1 - \rho^2{}_{(k,k+1)|(k-1)} \cdot B_{k|(k-1)}\right] \cdot \left[1 - \rho^2{}_{(k,j+1)|(k-1)} \cdot B_{k|(k-1)}\right]}} \qquad \textit{Equation 241}$$

which is also referred to as the partial correlation between X_k and X_j.

In summary, the PCM involves recursive calculations of two variables, the conditional fractile and conditional correlation coefficient. The recursion is performed in an optimized way within the correlation matrix $\boldsymbol{\Sigma}$ similar to Gaussian elimination. For more details on this method please see [73] and [74]. For algorithmic details, the MATLAB® code for the PCM method is provided in [74].

Generating Correlated Random Numbers

In RMC-TotalRisk, the dependency between system components is modeled by assigning correlation between the component hazard functions. For example, if there are two levees in a system on the same river reach, the flood hazards at each levee will be strongly positively correlated. The system risk calculations require that the nonexceedance probabilities of the component hazard levels be generated from the MVN. The following steps are followed:

1. Create a vector of independent random numbers $\boldsymbol{r} = \{r_1, \cdots, r_k\}$, where $r_i \sim U[0,1]$
2. Convert the vector to standard Normal z-variates $\boldsymbol{z} = \{\Phi^{-1}(r_1), \cdots, \Phi^{-1}(r_k)\}$
3. Generate a correlated vector of z-variates $\boldsymbol{u} = \boldsymbol{A}\boldsymbol{z}$, where $\boldsymbol{A}$ is the Cholesky decomposition of the correlation matrix $\boldsymbol{\Sigma}$.
4. Convert the vector back to random numbers $\boldsymbol{r} = \{\Phi(u_1), \cdots, \Phi(u_k)\}$, where $r_i \sim U[0,1]$

Appendix G – Sensitivity Analysis

Sensitivity analysis quantifies variations in the model outputs caused by possible variations in model inputs [75]. It can be used to rank the relative importance of different model inputs and to determine which model inputs contribute the most to the variance of model outputs.

In RMC-TotalRisk, a derivative-based sensitivity analysis is provided for the Event Tree and Risk Analysis components of the software. The partial derivative measures how sensitive an output component f is with respect to an input parameter θ_i when all other input parameters are held fixed.

$$\frac{\partial f}{\partial \theta_i} \ (i = 1,2,\cdots,n) \qquad \textit{Equation 242}$$

In RMC-TotalRisk, the partial derivatives are evaluated using numerical differentiation with the two-point formula:

$$\frac{\partial f}{\partial \theta} = \frac{f(\theta + h) - f(\theta - h)}{2h} \qquad \textit{Equation 243}$$

where h represents a small change in θ. To compare the sensitivities of different state variables with respect to different parameters, the dimensionless sensitivity, or *sensitivity index*, of an output component f to an input can be defined by normalizing with the mean of the θ_i and f:

$$SI_i = \frac{\mu_{\theta_i}}{\mu_f}\frac{\partial f}{\partial \theta_i} \ (i = 1,2,\cdots,n) \qquad \textit{Equation 244}$$

where μ_θ is the mean value for each input; and μ_f is the mean value of each output. Alternatively, the sensitivity index can be defined by normalizing by the variance of the θ_i and f. The variance of a single model output component f is given by:

$$\sigma_f^2 = \sum_{i=1}^{n}\sum_{1=j}^{n}\frac{\partial f}{\partial \theta_i}\frac{\partial f}{\partial \theta_j}\,\boldsymbol{\Sigma}(i,j) \qquad \textit{Equation 245}$$

where $\boldsymbol{\Sigma}$ is the covariance matrix for the input parameters. The variance of each parameter is in the diagonal of $\boldsymbol{\Sigma}$. For example, if the model has two inputs, Equation 245 can be written longhand as follows:

$$\sigma_f^2 = \left(\frac{\partial f}{\partial \theta_1}\right)^2 \sigma_{\theta_1}^2 + \left(\frac{\partial f}{\partial \theta_2}\right)^2 \sigma_{\theta_2}^2 + 2\left(\frac{\partial f}{\partial \theta_1}\right)\left(\frac{\partial f}{\partial \theta_2}\right)\mathrm{cov}(\theta_1,\theta_2) \qquad \textit{Equation 246}$$

where cov is the covariance between the pairwise input parameters. If the model inputs are independent and have no covariance, then Equation 246 reduces to:

$$\sigma_f^2 = \left(\frac{\partial f}{\partial \theta_1}\right)^2 \sigma_{\theta_1}^2 + \left(\frac{\partial f}{\partial \theta_2}\right)^2 \sigma_{\theta_2}^2 \qquad \textit{Equation 247}$$

The variance-based sensitivity index is then computed as:

$$SI_i = \frac{\sigma_{\theta_i}^2}{\sigma_f^2}\left(\frac{\partial f}{\partial \theta_i}\right)^2 \ (i = 1,2,\cdots,n) \qquad \textit{Equation 248}$$

Equation 248 is often referred to as the *main effect index* or the *contribution to variance* [76]. This provides the fractional contribution of variance from the input to the total output variance.

In the risk analysis, it is challenging to compute the partial derivatives of the output with respect to each input. Instead, the sensitivity index is derived from a linear regression of inputs and output from a Monte Carlo simulation. For each Monte Carlo realization, the sampled inputs θ and the resulting output y are stored in a matrix. Then, a multiple linear regression is estimated as:

$$y = \sum_{i=1}^{n} \beta_i \theta_i + \varepsilon \qquad \textit{Equation 249}$$

where the regression coefficient β_i measures the effect that input θ_i has on the predicted value y; and ε is the model error, or residual. The variance in θ_i and y can be estimated directly from the array of Monte Carlo samples. Then, the sensitivity index is computed as:

$$SI_i = \frac{\sigma_{\theta_i}^2}{\sigma_y^2} \beta_i^2 \ (i = 1,2,\cdots,n) \qquad \textit{Equation 250}$$

Correlation Measures

RMC-TotalRisk also provides correlation measures to describe the strength and direction of an association between the simulated input and output variables in the Event Tree and Risk Analysis. Pearson's correlation coefficient is a measure of linear correlation between two variables and is defined as:

$$\rho = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \qquad \textit{Equation 251}$$

Spearman's rank order correlation is estimated by first replacing the two variables by their ranks. Then, the Spearman correlation coefficient is calculated by estimating the Pearson correlation with the ranks instead of the values. Spearman's correlation is preferred for data that follow curvilinear, monotonic relationships.

Appendix H – Equivalent Annual Consequences

Equivalent Annual Consequences (EqAC) are applicable when considering changes in hazard, response, or consequences over the life of a project by discounting the future expected value to the base year resulting in an equivalent value over the life of the project. EqAC is a primary decision metric for planning in the USACE [10]. The methods applied in RMC-TotalRisk to calculate EqAC are based on procedures from the HEC-FDA User's Manual [13] as follows: Future year damage values are linearly interpolated between the base and most likely future year conditions and assumed constant from the most likely future year to the end of the analysis period. The analysis period (project life) is the period of time over which the plan has significant beneficial or adverse effects. It is normally 50 years and is not to exceed 100 years. Data requirements in RMC-TotalRisk to calculate EqAC are base year risk analysis and year, future year risk analysis and year, discount rate, and period of analysis. EqAC is computed as follows:

$$EqAC = CRF \cdot TPV \qquad \textit{Equation 252}$$

where $EqAC$ is the Equivalent Annual Consequence Value; CRF is the Capital Recovery Factor; and TPV is the Total Present Value. The capital recovery factor (CRF) is the ratio used to calculate the present value of an annuity (consequence). It is related to the annuity formula. The capital recovery factor can be represented as follows:

$$CRF = \frac{r \cdot (1+r)^n}{(1+r)^n - 1} \qquad \textit{Equation 253}$$

where r is the discount rate; and n is the planning period in years. The Total Present Value (TPV) is the sum of present values calculated for each year over the lifetime of the project:

$$TPV = \sum_{i=b}^{b+n} PV_i \qquad \textit{Equation 254}$$

where PV_i is the present value for a given year i; b is the base year; and n is the period of analysis. Present value (PV) is the value as of the base year of some amount in the future. Starting from the base year and extending to the period of analysis, the present value is estimated as follows:

$$PV_i = FV_i \cdot (1+r)^{-(i-b)} \qquad \textit{Equation 255}$$

where PV_i is the present value for a given year i; FV_i is the future value for the given year; r is the discount rate; and $(i-b)$ is the period of analysis represented as the given year minus the base year. The future value (FV) is calculated as a stepwise function of either linearly interpolating between base and future years or taking the future year expected annual consequences:

$$FV_i = \begin{cases} y_f, & i \geq y_f \\ EAC_b + \left(EAC_f - EAC_b\right)\left(\dfrac{i - y_b}{y_f - y_b}\right), & i < y_f \end{cases} \qquad \textit{Equation 256}$$

where FV_i is the value for any future year i; EAC_b is the expected annual consequence (damages) for the base year conditions b; and EAC_f is the expected annual consequence for the future year conditions f; y_f is the year of the future EAC_f estimate; y_b is the year of the base EAC_b estimate;.

Appendix I – National Flood Insurance Program

The purpose of constructing a levee is to reduce flood risk for the area behind the levee. Certified levees which reduce flood risk from the 0.01 annual exceedance probability (AEP) event, or base flood, are eligible for accreditation in the National Flood Insurance Program (NFIP) administered by the Federal Emergency Management Agency (FEMA). FEMA can only accredit certified levees for its flood hazard and risk mapping effort which meet minimum design, performance, operation, and maintenance standards that are consistent with criteria specified in Title 44, Part 65.10 of the Code of Federal Regulations or CFR [77].

The USACE plays a significant role in managing flood risk nationwide since the enactment of the Flood Control Act of 1917 including design, construction, and risk assessment of levee systems [2]. The USACE uses a risk-informed approach to perform NFIP Levee System Evaluations (LSEs) to make a recommendation about whether to certify and accredit a levee system.

USACE guidance for conducting LSEs to assess levee accreditation is currently outlined in Engineering and Construction Bulletin (ECB) 2019-11 [78]. The ECB specifies that USACE and FEMA should rely on Semi-Quantitative Risk Assessments (SQRAs) or Quantitative Risk Assessments (QRAs) to make NFIP accreditation recommendations [78].

All QRAs must include a computation of assurance of the 0.01 AEP, which is the probability that the 0.01 AEP event will not be exceeded [78]. The 0.01 AEP is the annual probability that the leveed area will be inundated due to levee overtopping or breach in any given year. If the computed assurance of the 0.01 AEP is less than 65%, then the recommendation should be *Do Not Accredit* the levee system. Substantial evidence with high confidence and limited uncertainty is needed to support a NFIP levee accreditation recommendation when the assurance is less than 65%. If the computed assurance for the 0.01 AEP is greater than 85%, then the recommendation should be to *Accredit* the levee system. Only strong evidence or significant uncertainty should result in a recommendation of *Do Not Accredit* if the assurance is greater than 85%. If the computed assurance of containing the 0.01 AEP event falls between 65% and 85% then the accreditation recommendation must be supported based on uncertainty, past system performance, and other factors [78].

Annual Probability of Inundation

Computation of assurance of the 0.01 AEP for the NFIP requires an estimate of the annual probability of inundation (API), which is the probability that the leveed area will be inundated due to levee overtopping or breach in any given year. To compute the API, the user must select the *Profile Hazard Type* and specify the *Hazard Threshold* for the desired system component as shown in Figure 77:

- **Profile Hazard Type**: The hazard type used for constructing risk profiles and used for estimating the probability of exceeding the hazard threshold. For a levee accreditation analysis, the hazard type should be stage or water surface elevation.

- **Hazard Threshold**: The probability of hazard levels exceeding the threshold will be recorded in the risk simulation. For a levee accreditation analysis, the hazard threshold should be the top of levee height or elevation.

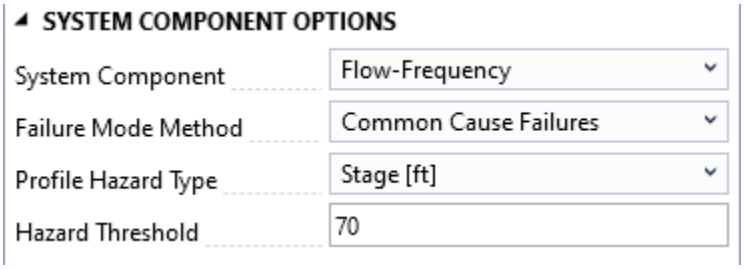


Figure 77 - Setting the profile hazard type and threshold level for the system component.

For a typical levee, the leveed area can be inundated in three ways: 1) the levee fails prior to overtopping; 2) the levee overtops and fails; and 3) the levee overtops without failing. Therefore, the probability of inundation is the union of three events:

$$API = P(\overline{OT} \cap Fail) \cup P(OT \cap Fail) \cup P\left(OT \cap \overline{Fail}\right) \quad \textit{Equation 257}$$

where $P(\overline{OT} \cap Fail)$ is the probability that the levee fails prior to overtopping; $P(OT \cap Fail)$ is the probability that the levee overtops and fails; and $P\left(OT \cap \overline{Fail}\right)$ is the probability that the levee overtops without failing. The first two events are equivalent to the total probability of failure for the levee, which in RMC-TotalRisk is computed as:

$$P(\overline{OT} \cap Fail) \cup P(OT \cap Fail) = \sum_{i=1}^{n} P(x_i) \cdot P(F|x_i) = \alpha_{\mathrm{F}} \quad \textit{Equation 258}$$

where $P(x_i)$ is the probability of the hazard level x_i; $P(F|x_i)$ is the conditional probability of failure given the hazard level x_i; and α_{F} is the probability of failure from all failure modes. The third event, which is the probability the levee overtops without failing, is computed as:

$$P\left(OT \cap \overline{Fail}\right) = P(x \geq x_T) = \sum_{i=x_T}^{n} P(x_i) \cdot \{1 - P(F|x_i)\} \quad \textit{Equation 259}$$

where x_T is the hazard threshold set by the user, which in this case is the top of levee height; $\{1 - P(F|x_i)\}$ is the probability of non-failure given the hazard level x_i, which is simply the complement of the probability of failure at a given hazard level. The probability of inundation is then computed as follows:

$$API = \sum_{i=1}^{n} P(x_i) \cdot P(F|x_i) + \sum_{i=x_T}^{n} P(x_i) \cdot \{1 - P(F|x_i)\} \quad \textit{Equation 260}$$

For NFIP applications, the hazard threshold x_T should be the top of levee height, or elevation. The hazard threshold defines when overtopping (OT) occurs at the levee. However, this formula is generalized in RMC-TotalRisk, so the calculation will work for other threshold-based applications as well.

Assurance

The level of assurance (e.g., the levee can be accredited with 85% assurance) is synonymous with the confidence level, i.e., the leveed area is inundated with a probability of 0.01 or less, with 85% confidence. To assess the level of assurance with RMC-TotalRisk, the risk analysis must be simulated with full uncertainty. The user must enter the number of Monte Carlo realizations. The default is 1,000 to ensure reasonably accurate confidence intervals. For accurate assurance calculations, it is recommended to run 10,000 realizations.

After the risk analysis is complete, the user can navigate to the Diagnostics tab, and then the Assurance tab. Figure 78 shows the standard cumulative distribution plot for assurance. The API is on the x-axis, and the non-exceedance probability of the uncertainty in the API is on the y-axis. For this example, the mean API is shown as a vertical, dashed blue line at 0.00643. The target level of 0.01 (100-yr), which is shown as a vertical orange line, is shown to have an assurance level of 84.6%. The API and assurance results can also be plotted as a kernel density as shown in Figure 79.

If the assurance level is less than 65%, the target level line will be plotted as a vertical red line. If the assurance is between 65% and 85%, the target level line will be plotted as an orange line. If the assurance is greater than 85%, it will be plotted as a greed line. In the example below, the assurance of containing the 0.01 AEP event falls between 65% and 85%. Therefore, the levee accreditation recommendation must be supported based on uncertainty, past system performance, and other factors [78].

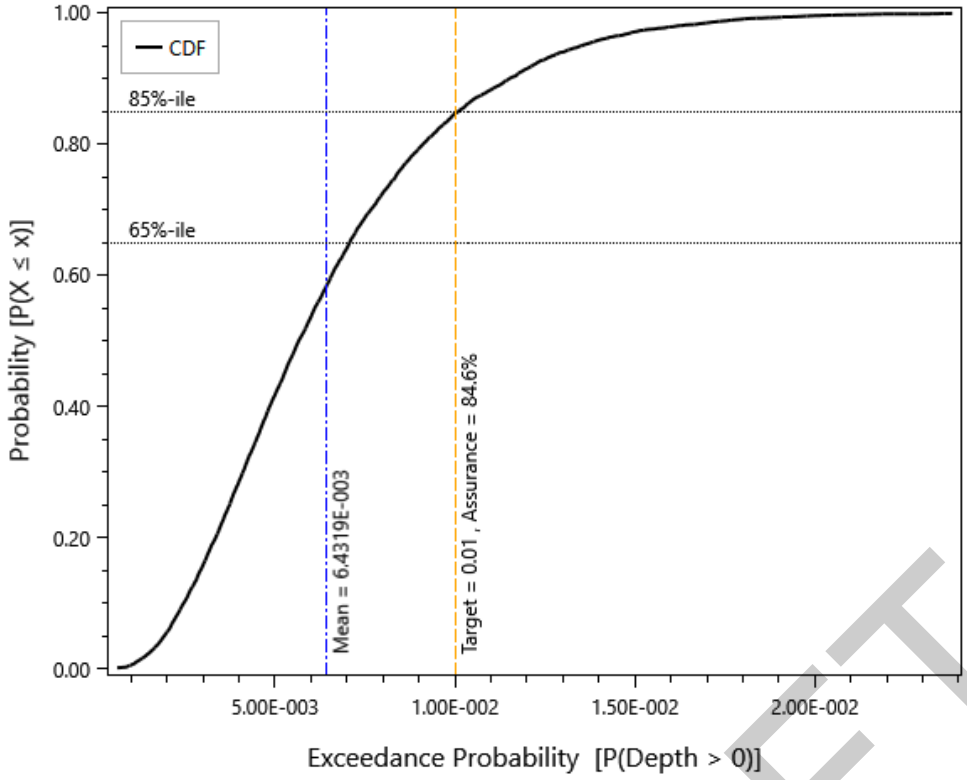


Figure 78 - Example of the cumulative distribution plot for NFIP assurance.

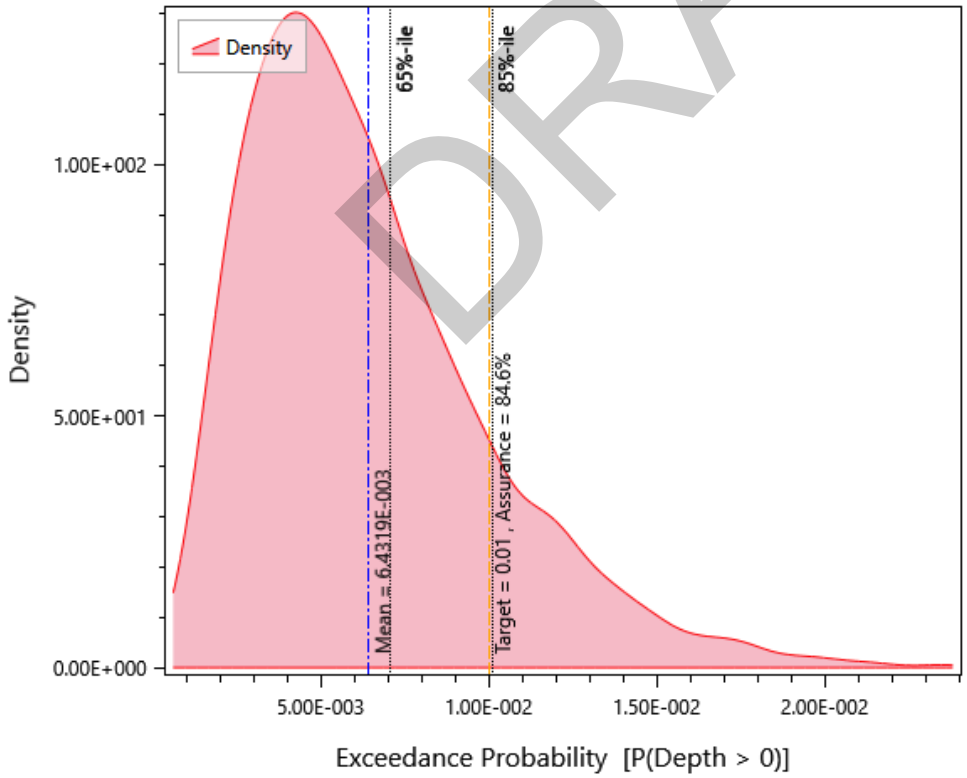


Figure 79 - Example of the density plot for NFIP assurance.

Additional Risk Measures

The "probability of inundation" measure computed for levee accreditation is just a flood hazard measure and is not a true risk measure. There is no consideration for consequences, just the probability of an inundation depth being greater than zero. RMC-TotalRisk provides some additional true risk measures that can be useful for risk-based design purposes for levees and dams. The following subsections briefly describe these measures and how they could be used to support risk-informed investment decisions. More details on the risk measures are provided in the **Risk Measures** section of the main report.

Consequence Threshold Probability

The user can enter a consequence threshold n. The probability of consequences exceeding the threshold will be recorded in the risk simulation. The default consequence threshold is 0. The exceedance probability is interpolated from the F-N curve, which is a survival function, for each risk type:

$$S(n) = 1 - F(n) = P(N \geq n) \qquad \textit{Equation 261}$$

After simulating risk with full uncertainty, a kernel density plot and summary statistics are provided (see the **Diagnostics** section). This risk measure provides the expected probability of exceedance of a specified consequence level. When designing a new levee, the objective could be that a consequence threshold of $1M should only be exceeded 1:100 years, i.e., a 0.01 annual exceedance probability. This risk measure can be used to determine if that design objective has been satisfied with a desired level of assurance.

Value-at-Risk

The user can enter an exceedance probability α, such as 0.01. This exceedance probability is used for computing the Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR). The default exceedance probability is 0.01. The VaR provides the minimum consequences for the user-specified exceedance probability α. The value is interpolated from the F-N curve for each risk type:

$$S^{-1}(\alpha) = n \qquad \textit{Equation 262}$$

The objective for a new levee could be that the minimum consequences allowable for a 100-year flood must be $1M. This risk measure can be used to determine if that design objective has been satisfied with a desired level of assurance.

Conditional Value-at-Risk

The Conditional Value-at-Risk (CVaR) provides the mean (or expected) consequences for the user-specified exceedance probability α. The value is computed by integrating under the F-N curve for each risk type in the prescribed probability range:

$$\mathbb{E}[N|N \geq \beta] = \frac{1}{\alpha}\int_{\beta}^{\infty} x \cdot dS(x) \qquad \textit{Equation 263}$$

where $\beta = S^{-1}(\alpha) = n$. CVaR is estimated using numerical integration with adaptive Simpson's rule (see **Appendix D** for more details).

The objective for a new levee could be that the mean consequences allowable for a 100-year flood must be less than $10M. This risk measure can be used to determine if that design objective has been satisfied with a desired level of assurance.