### Uncertainty in Risk Assessments: Concepts and Principles

#### Summary of Uncertainty Types: Terminology and a Comparison from Selected Sources

<table>
<thead>
<tr>
<th>Types</th>
<th>PRA</th>
<th>Reliability Engineering</th>
<th>Risk Engineering</th>
<th>PRA</th>
<th>PRA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Physical Variability</td>
<td>Non-Cognitive (Quantitative) Uncertainty</td>
<td>Inherent Uncertainty</td>
<td>Inherent Uncertainty</td>
<td>Aleatory Uncertainty (Stochastic Uncertainty or Randomness)</td>
</tr>
<tr>
<td>2</td>
<td>Parameter Uncertainty</td>
<td>Statistical Uncertainty</td>
<td>Statistical Uncertainty</td>
<td>Parameter Uncertainty</td>
<td>Parameter Uncertainty</td>
</tr>
<tr>
<td>3</td>
<td>Modeling Uncertainty</td>
<td>Modeling Uncertainty</td>
<td>Modeling Uncertainty</td>
<td>Model Uncertainty</td>
<td>Model Uncertainty</td>
</tr>
<tr>
<td>4</td>
<td>Completeness Uncertainty</td>
<td>Cognitive (Qualitative) Sources of Uncertainty</td>
<td>Human Error</td>
<td>“Completeness” is a requirement and not a specific type of uncertainty</td>
<td>Epistemic Uncertainty</td>
</tr>
</tbody>
</table>

#### References and Foot Notes (Selected excerpts from References 1, 4, and 5 start on page 2):

6 - As per reference 1 (p. 315), these three types of uncertainty are from the IEEE/ANS Procedures Guide (NUREG/CR-2300, 1983).
7 - As per reference 2 (p. 5), cognitive and qualitative sources of uncertainty relate to vagueness in definitions of certain parameters, definitions of interrelationships among the parameters especially for complex systems, and human factors.
8 - As per reference 4, “completeness” does not miss a dominant scenario (p. 5), is a requirement (p. 27), develops the scenario model in a hierarchical fashion (p. 27), implies a significant effort in the development of the scenario set (p. 45), and includes uncertainty in data gathering and database development (pp.184 & 185).
Uncertainty in Risk Assessments: Concepts and Principles


1. **Risk** is a measure of the potential loss occurring due to natural or human activities (p. 1).
2. **Risk analysis** is the process of characterizing, managing, and informing others about existence, nature, magnitude, prevalence, contributing factors, and uncertainties of the potential losses (p. 1).
3. **Risk assessment** is a formal and systematic analysis to identify or quantify frequencies or probabilities and magnitude of losses to recipients due to exposure to hazards (physical, chemical, or microbial agents) from failures involving natural events and failures of hardware, software, and human systems. Generally speaking, a risk assessment amounts to addressing three basic questions proposed by Kaplan and Garrick (p. 7):
   a. What can go wrong?
   b. How likely is it?
   c. What are the losses (consequences)?
4. **Uncertainty** is a measure of the “goodness” of an estimate...Without such a measure, it is impossible to judge how closely the estimated value relates to or represents reality. Uncertainty arises from lack of or insufficient knowledge (p. 197).
5. The simplest way to represent the probability as a measure of uncertainty is to use the mean and variance...and any epistemic uncertainty associated with these estimations. Estimations of the mean and variance are themselves subject to uncertainty. In a classical [statistics] estimation approach only the confidence intervals of the mean and variance can be estimated (p. 207).
6. **Note:** Chapter 5, Uncertainty Analysis: Addresses uncertainty types, measures of uncertainty, uncertainty propagation, and graphical representation of uncertainty.


1. The objective of a probabilistic risk analysis (PRA) is the quantification of risk from made man-made and natural activities (p. 313).
2. Two major types of uncertainty need to be differentiated in a PRA: (1) Uncertainty due to physical variability and (2) Uncertainty due to lack of knowledge (p. 314).
3. As we gain more knowledge, knowledge uncertainty will decrease, however, physical variability will not decrease. We will know the variable behavior better and be able to quantify it more precisely; however, the variability itself will not diminish (p. 314).
4. Knowledge uncertainty consists not only of imprecision in parameter estimates, but also incompleteness in modeling, vagueness in appropriate engineering estimates, indefiniteness in the applicability of the model, and doubtfulness and vagueness in the interpretability of results produced by a model. Knowledge uncertainty is generally more difficult to quantify than physical variability (p. 314).
5. Knowledge uncertainties are quantified in PRAs using either uncertainty analyses or sensitivity analyses (p. 314).
6. **Sensitivity analyses** are straightforward conceptually and consist of determining changes in outputs which result from changes in inputs (p. 314). Uncertainties which are due to alternative models or alternative assumptions need to be separately considered by sensitivity analyses. Sensitivity analysis is not performed by assigning uncertainties to parameters and propagating these uncertainties. In specific cases, the effects of different modeling assumptions can be as large as uncertainties on parameters (pp. 320 & 321).
7. **Uncertainty analyses** in PRAs utilize Bayesian approaches, probabilistic approaches (similar to the Bayesian approach except the distribution assigned to the parameter represents the physical variations from plant to plant), and sometimes classical statistical approaches (p. 314).
Uncertainty in Risk Assessments: Concepts and Principles


9. Parameter uncertainties include not only imprecisions due to small samples of recorded data, but also uncertainties in experts’ judgments of parameter values when there are not recorded data (p. 315).

10. The selected shapes of the prior distributions [when the Bayesian or probabilistic approaches are used] vary from PRA to PRA; log-normal distributions are often used based on the WASH 1400 [NUREG-75/014, Oct 1975] precedent but so are gamma distributions, log-uniform distributions, and discrete probability distributions with generally less than 10 points (p. 317).

11. [When the uncertainty distribution for the parameter is characterized by the log-normal distribution], uncertainties on parameter estimates are expressed as error factors. The error factor is the ratio of the upper 95th percentile to the median or 50th percentile. [Error factor, k, is directly proportional to the number of standard deviations in log units from the logarithmic mean since error factor can be expressed as \( k = \exp(1.645*\sigma) \) where 1.645 is the 95th percentile for \( \ln \lambda \) and \( \lambda \) is the random variable.] Error factors range from 1.3 to 30. The smallest factors (1.3 to 2) apply to the higher event frequencies (1 or more per reactor year) for which a larger amount of recorded data generally exists. Accident frequencies in the vicinity of \( 1 \times 10^{-7} \) per reactor year generally have resulting error factors in the range of 20-30. Thus, the larger error factors are generally associated with the smaller frequency estimates (p. 318).

12. Modeling uncertainty can be divided into two subcategories: (1) Indefiniteness in the model’s comprehensiveness (i.e., does the model account for all the variables which can significantly affect the results), and (2) Indefiniteness in the model’s characterization (i.e., refers to the uncertainties in the relations and descriptions used in the model). Even if the pertinent variables are included in the model, appropriate relationships among the variables may not be described (p. 316).

13. Completeness uncertainties are the uncertainties as to whether all the significant phenomena and all the significant relationships have been considered in the PRA. Completeness uncertainties are similar in nature to modeling uncertainties but occur at the initial stage in the PRA. There are two subcategories of completeness uncertainties: (1) Contributor uncertainties (i.e., uncertainty as to whether all the pertinent risks and all the important accidents have been included) and (2) Relationship uncertainties (i.e., uncertainty as to whether all the significant relationships are identified which exist among the contributors and variables) (p. 317).


1. PRA models are constructed as probabilistic models to reflect the random nature of the constituent basic events such as initiating events and component failures. Randomness is one manifestation of a form of uncertainty that has come to be called aleatory uncertainty. Therefore, a PRA is a probabilistic model that characterizes the aleatory uncertainty associated with accidents at nuclear power plants (NPPs). The focus of this document is guidance in dealing with another area of uncertainty, namely epistemic uncertainty. This uncertainty is associated with the incompleteness in the analysts’ state of knowledge about accidents at NPPs and has an impact on the results of the PRA model (p. 28).

2. Types of epistemic uncertainty:
   a. Parameter uncertainty is the uncertainty in the values of the parameters of a model given that the mathematical form of that model has been agreed to be appropriate (p. 30).

170511 Tim.Adams@NASA.gov
b. **Model uncertainty** is related to an issue for which no consensus approach or model exists and where the choice of approach or model is known to have an effect on the PRA model (p. 30).

c. **Completeness uncertainty** relates to risk contributors that are not in the PRA model. These types of uncertainties either are ones that are known but not included in the PRA model or ones that are not known and therefore not in the PRA model (p. 15). Lack of completeness is not in itself an uncertainty, but recognition of the limitations in the scope of the model. However, the result is an uncertainty about where the true risk lies. This uncertainty also can be thought of as a type of model uncertainty. However, completeness uncertainty is discussed separately because it represents a type of uncertainty that cannot be quantified and because it represents those aspects of the system that are, either knowingly or unknowingly, not addressed in the model (p. 30).

3. Assessing the impact of uncertainty:

a. Although many sources of uncertainty may exist in a PRA, they do not all have an impact on a particular decision. Those sources that can affect the results used to support a decision are identified as **key sources of uncertainty**, and these demand attention. The way that the uncertainties are assessed is a function of the way the acceptance criteria (or acceptance guidelines) for the decision are defined and the way that the uncertainties are characterized with regard to the acceptance guidelines (p. 31).

b. **Acceptance guidelines** are defined to demonstrate, with reasonable assurance, that the risk change is small. This decision has to be based on a full understanding of the contributors to the PRA results and the impacts of both the uncertainties that are explicitly accounted for in the results and those uncertainties that are not. The term “acceptance guidelines” rather than “acceptance criteria” is used primarily to recognize that the numerical results do not capture all of the analysis uncertainty (p. 31).

c. **Parameter uncertainty**: The appropriate measure for comparison is the mean value of the uncertainty distribution on the corresponding metric. The primary issue with parameter uncertainty is its effect on the calculation of the mean, and specifically, on the relevance and significance of the state-of-knowledge correlation (p. 31). The uncertainties on the input parameters need to be combined in an appropriate manner to provide an assessment of this type of uncertainty on the PRA results. An important aspect of this propagation is the need to account for what has been called the **state-of-knowledge correlation (SOKC)** or **epistemic correlation**. This concern arises when the same parameter (including its uncertainty) is used to quantify the probabilities of two or more basic events. Most of the PRA software in current use has the capability to propagate parameter uncertainty through the analysis, taking into account the SOKC to calculate the probability distribution for the results of the PRA. In some cases, however, it may not be necessary to consider the SOKC (p. 14).

d. **Model uncertainty**: The typical response to a modeling uncertainty is to choose a specific modeling approach to be used in developing the PRA model. Although it is possible to embed a characterization of model uncertainty into the PRA model by including several alternate models and providing weights (probabilities) to represent the degree of credibility of the individual models, this approach is not usual. In dealing with model uncertainties, it is helpful to identify whether the model uncertainties can alter the logic structure of the PRA model or whether their primary impact is on the probabilities of the basic events of the logic model. One approach to dealing with a specific model uncertainty is to adopt a **consensus model** that essentially removes the uncertainty related to the choice of model from having to be addressed in the decision making. What is removed by adopting a consensus model is the need to consider other models as alternatives (pp. 31 & 32).

e. **Completeness Uncertainty**: The issue of completeness of scope and level of detail of a PRA can be addressed for those scope and level of detail items for which methods are available and, therefore, some understanding of the contribution to risk exists. The true unknowns (i.e., those related to issues whose existence is not recognized) cannot be addressed analytically. The principles of safety margins and defense in depth play a critical role in addressing this source of uncertainty (p. 33).
Uncertainty in Risk Assessments: Concepts and Principles


1. The first step in doing a PRA is to structure the problem, which means to build a model for the physical situation at hand. This model is referred to as the model of the world. It may occasionally be referred to it as the “model” or the “mathematical model.” It is built on a number of model assumptions and typically includes a number of parameters whose numerical values are assumed to be known (p. 96).

2. There are two types of models of the world, deterministic and probabilistic (p. 96).

3. The uncertainty described by the model of the world is sometimes referred to as “randomness” or “stochastic uncertainty.” Stochastic models of the world have also been called aleatory models. It is important to point out that models of the world, regardless of whether they are deterministic or aleatory, deal with observable quantities [e.g., distance and time] (p. 97).

4. Each model of the world is conditional on the validity of its assumptions and on the availability of numerical values for its parameters. The epistemic model represents the state of knowledge regarding the numerical values of the parameters and the validity of the model assumptions (p. 97). Unlike aleatory models of the world, the epistemic models deal with non-observable quantities. Failure rates and model assumptions are not observable quantities (p. 98).

5. The issue of alternative model assumptions is usually handled by performing sensitivity studies (p. 97).

6. Regardless if the probability is for an aleatory or epistemic model, the concept of probability remains the same. However, there are two major interpretations of probability: (1) Classical or frequentist which is based on a limit of relative frequencies and (2) Bayesian which is based on a measure of degree of belief (pp. 98 and 99).

7. If one adopts the frequentist interpretation of probability, then one is not allowed to use epistemic models. The numerical values of the parameters of the model of the world must be based on statistical evidence only. A number of methods have been developed for producing these numerical values. A widely used method for producing point estimates of the parameters is the method of maximum likelihood. As the statistical evidence becomes stronger, the Bayesian posterior distribution will tend to have a mean value that is equal to the maximum likelihood estimates (MLE). In other words, any prior beliefs will be overwhelmed by the statistical evidence (p. 115).

8. The lognormal distribution is used frequently in safety studies as the epistemic distribution of failure rates (λ) and uses the Error Factor as a measure of dispersion. The lognormal distribution is skewed to the right. This (in addition to its analytical simplicity) is one of the reasons why it is chosen often as an epistemic distribution for failure rates. It allows high values of λ that may represent extreme environments (pp. 102 & 103).

9. When evidence becomes available, it is natural to change the epistemic models to reflect this new knowledge. The typical problem encountered in practice involves an aleatory model of the world. The evidence is in the form of statistical observations. The analytical tool for changing (“updating”) our epistemic distributions of the parameters of the aleatory model is Bayes’ Theorem. Direct assessments of model parameters, like direct assessments of failure rates, should be avoided, because model parameters are not directly observable (they are “fictional”). The same observation applies to moments of distributions, for example, the mean and standard deviation (pp. 103, 114, & 124).

10. Intuitive estimates of the mode or median of a distribution have been found to be fairly accurate, whereas estimates of the mean tend to be biased toward the median. This has led to the suggestion that “best” estimates or “recommended” values, which are often offered by engineers, be used as medians. In assessing rare-event frequencies, however, the possibility of a systematic underestimation or overestimation [“displacement bias”], even of the median, is very real (p. 114).
Uncertainty in Risk Assessments: Concepts and Principles

11. In regards to a major risk study of nuclear power plants, objections were raised both to the use of expert opinions (with complaints that voting is replacing experimentation and hard science) and to the process of using expert opinions (for example, the selection of the experts). The latter criticism falls outside the mathematical theory…however, the view that voting replaces hard science is misguided. The (epistemic) probabilities of models are an essential part of the decision-making process. Unfortunately, many decisions cannot wait until such evidence becomes available, and assessing the model probabilities from expert opinions is a necessity. (Incidentally, such an assessment may lead to the decision to do nothing until experiments are conducted) (p. 114).

12. In regards to the uncertainties in the numerical values of the parameters of a given model (parameter uncertainty) rather than on the uncertainty regarding the validity of the model itself, [typically] the simulation technique is used to propagate the uncertainty in the risk model outputs induced by epistemic uncertainties in input parameter values. With uncertainty propagation bounding values to output functions of the PRA can be estimated. These bounds are associated with a probability that bounds include the true value of the numerical result predicted by the model. The term “output function” [response variable] refers to the function corresponding to the risk metric of interest. The term “input parameters” [one or more random variables] represents the uncertain input to the output function. (p. 206). The uncertainty associated with the risk model assumptions is handled with sensitivity analysis (p. 207).

13. Techniques for propagation of uncertainties (p. 207):
   a. Simulation: The distributions for input parameters are mapped using crude Monte Carlo or Latin Hypercube sampling (LHS) technique to obtain an empirical distribution for the output function. In the NASA guide, only the simulation technique is described because this technique has become the industry standard for propagating uncertainties;
   b. Moment propagation: First and second moments of the input parameters are mapped to obtain mean and variance of the output function using variance/covariance propagation; and
   c. Discrete Probability Distribution: The distributions for input parameters are converted to discrete probability distribution before mapping. The resulting distribution for the output function is empirical.

14. The process of selecting the set of possible values for each parameter consistent with its distribution is called sampling (p. 208).
   a. Crude Monte Carlo Sampling: Sampling is completely random. That is, a value is drawn at random from the distribution for each input parameter. Of course, samples are more likely to be drawn in the areas of distribution where the probability of occurrence is higher. Because of this property, low probability areas of the distribution (i.e., tail of the distribution) may not represented adequately in the samples. As a result, a relatively high number of iterations is required to obtain reliable estimates of the output function. This issue is particularly problematic for risk and reliability models that employ skewed probability distributions [e.g., lognormal distribution]. This problem led to development of the following technique (p. 210).
   b. Latin Hypercube Sampling (LHS): Sampling is based on “sampling without replacement.” In this technique the cumulative distribution function for an input parameter is divided up into intervals. A single value is sampled at random from within each interval (or stratification) according to the probability distribution of the input parameter. Because the coverage of sampling over the input domain is more uniform, a smaller number of samples is required. For this reason LHS is more appealing than crude Monte Carlo sampling (p. 210).

15. Achieving Convergence: The precision in the propagated distributions is improved by increasing the number of samples. It is important to run enough iterations so that the statistics generated for the output function are reliable. Therefore, care should be exercised to include enough sample iterations to achieve statistical regularity. By statistical regularity we mean as more iterations are run, the distribution for the output function becomes more stable as the statistics describing the distribution change less and less with additional iterations (p. 211).
It is important that epistemic dependency among variables of a risk model is correctly accounted for in the uncertainty propagation. Dependence among uncertain parameters is often specified in terms of correlation coefficient, which has a value between –1 and 1. Due to lack of data, the PRA models often assume that the dependent variables are fully correlated. **Note:** Suggested techniques are provided on the handling of correlation (pp. 217 & 218).

---

**End Notes**


2 – This assumes the likelihood of occurrence is a “well grounded probability.” When rare events cannot be adequately treated on the basis of measurable probabilities and traditional risk analysis, such risks are studied and treated from the perspective of **genuine uncertainty**. Genuine uncertainty can often not be reduced significantly by attempting to gain more information about the phenomena in question and their causes. (Ref. Ritchey, T, et al, *Modeling Society’s Capacity To Manage Extraordinary Events*, Society for Risk Conference in Paris, November 2004, p. 1, [http://www.swemorph.com/pdf/sra.pdf](http://www.swemorph.com/pdf/sra.pdf))