

**Methods for Probabilistic Uncertainty Analysis and Bayesian Analysis
with Examples of Statistically Analyzing Data to Revise MMOD Risk Estimates and Compare Models**

**Dr. William Vesely
Manager, Technical Risk Assessment
Safety Assurance and Requirements Division
Office of Safety and Mission Assurance**

August 23, 2017

Table of Contents

Why Quantify Uncertainty	4
Overview of the Technical Material Presented	4
Definition of Probabilistic Uncertainty Analysis	5
Steps Involved in Characterizing and Quantifying Uncertainties Probabilistically	6
Differentiating the Quantify to be Estimated Versus Its Estimated Value	7
Characterizations of an Uncertainty Distribution	7
Physical Uncertainty Versus Knowledge Uncertainty	8
Specific Characteristics of an Uncertainty Distribution	9
Estimating an Uncertainty Distribution for a Basic Quantity	10
<i>The Estimated Central Value</i>	10
<i>The Estimated Upper Bound Value</i>	10
<i>The Assigned Distribution Family</i>	11
<i>Beta Distribution</i>	11
<i>Uniform Distribution</i>	12
<i>Normal Distribution</i>	12
<i>Gamma Distribution</i>	13
<i>Lognormal Distribution</i>	14
Propagating Input Uncertainties in a Model	15
Common Ways of Handling Dependencies Among Uncertain Quantities	16
Considerations Involved in Quantifying the Uncertainty in an MMOD Risk Estimate	17
Handling of Alternative Models and Associated Modeling Uncertainty	19
Example of Handling Alternative Models in an MMOD Risk Estimate	23
Determining Uncertainty Contributions to an Overall Uncertainty	23
Separating Physical and Knowledge Uncertainty Contributions	25
Revising an Estimate and Its Uncertainty Using Observed Data	27
Revising a Risk Estimate with Zero Observed Failures Using a Poisson Model	30

Table of Contents Continued

Revising a Risk Estimate with an Observed Number of Events Using a Poisson Model	30
Using Fuzzy Observed Data	31
Comparing the Consistency of a Model Prediction with Observed Data	31
Using Observed Data to Compare Different Models	33
Uses of Quantified Uncertainties in Decision-Making	37
<i>Assessing Whether the Uncertainty Allows an Estimate to be Meaningfully Used</i>	38
<i>Determining the Confidence in the Satisfaction of a Requirement or Goal</i>	38
<i>Accounting for the Uncertainty in a Statistical-Decision Framework</i>	40
Examples of MMOD Risk Uncertainty Analysis and Data Analysis	43
<i>Comparison of Gamma Versus Lognormal Uncertainty Distributions</i>	43
<i>Comparing and Updating Model Predictions with Observed Data of No MMOD Occurrences</i>	46
<i>Comparing and Updating Model Predictions with Observed Data of an MMOD Occurrence</i>	50
References	53
Examples of WINBUGS Scripts for a Poisson Likelihood and Lognormal Uncertainties	55
Flow Charts for Designing Experiments to Update MMOD Risk Estimates with Observed Data	56

Why Quantify Uncertainty

Quantification of uncertainty provides a quantitative measure of the precision and accuracy of an estimate. The precision is a measure of the spread of the estimate. The accuracy measures the possible difference of the true value from the estimated value. The lack of accuracy is also called the bias in the estimate. Without quantification of the uncertainty in the estimate there is no definitive knowledge of how to interpret and use the estimate. With regard to a risk estimate, when the uncertainty is not quantified there is no definite knowledge as to how low or high the true risk value can be from the estimated value.

As importantly, quantification of the uncertainty provides the level of confidence in a requirement or goal being met by the estimate. By quantifying the uncertainty in an estimate, such as a risk estimate, the appropriate estimated value from the uncertainty characterization can be selected that provides the desired confidence that the requirement or goal is required. Without the quantification of uncertainty one does not have any definite confidence whether the requirement or goal is satisfied. Without the characterization and quantification of uncertainty one does not know how conservative or non-conservative an estimated value may be. This document describes approaches which can be used for systematically using quantified uncertainties in decision-making.

Characterization and quantification of the uncertainty also provides knowledge of whether the uncertainty is so large as to require a re-assessment to reduce the uncertainty so the estimate can be meaningfully used in decision-making. The quantification of uncertainty does not need to be exact and rarely can it be and is not necessary. Only the approximate size and key characteristics need to be determined as is true for all practical, quantitative uncertainty assessments. This document describes techniques that can be used and gives example applications.

Quantification of uncertainties also allows prioritization of the uncertainty contributors to the overall uncertainty in an estimated result. The impact of reducing one or more contributor uncertainties can also be determined showing where effort can be focused to optimally reduce the uncertainty in the result. And finally, and importantly, by quantifying uncertainty, observed data can be systematically used to update and revise an estimate and its associated uncertainty. The observed data does not need to be precise, can be both statistical and subjective, and can be sparse. This document provides techniques and examples.

Overview of the Technical Material Presented

This document describes techniques for carrying out and utilizing probabilistic uncertainty analysis. Formal developments of key relationships and key results are presented to provide a systematic, justifiable approach for characterizing and quantifying uncertainties in estimated quantities. Examples of such uncertain, estimated quantities include risk estimates, cost estimates, and lifetime estimates. Often uncertainties in estimated quantities are treated with ad hoc approaches or with a slight-of-hand pronouncements. Worse yet, often uncertainties are ignored. To systematically treat uncertainties, approaches specifically developed to treat uncertainties are required where the uncertainties include both data and model uncertainties. Also, in many cases and particularly in risk estimation, only sparse data are available and often it is fuzzy. Specific approaches are needed which are capable of systematically analyzing this data. Finally, specific approaches are needed to systematically handle

uncertainties in decision-making especially when the uncertainties can be large. The family of approaches which systematically deal with uncertainties is probabilistic uncertainty analysis which is the subject of this document.

Probabilistic uncertainty analysis quantifies uncertainties using probabilistic approaches. It is the standard technique used in reliability and risk assessments for quantifying uncertainties. It is also the standard technique used in NASA probabilistic risk assessments (PRAs). This document was prepared to provide specific methods and techniques for carrying out and applying probabilistic uncertainty analysis in broader applications. The methods and techniques are extracted from various sources to provide one accessible source. Using basic probability and statistical principles, the formulations and rationale are given for the derived procedures and results. Uncertainties can only be meaningfully quantified using probability and statistical principles and it is important to understand their role in the procedures and results that are given in the document.

The examples in the document focus on characterizing and quantifying uncertainties associated with the estimate of Micrometeoroid and Orbital Debris (MMOD) risk to a spacecraft. However the methods and techniques are generally applicable. Probabilistic uncertainty analysis has the capability to account for all types of uncertainty including assumption and modeling uncertainties as well as statistical data uncertainties. Observed data can furthermore be systematically analyzed to statistically revise an estimate and its uncertainty where the observed data can be sparse and be fuzzy. Probabilistic uncertainty analysis has the added feature of allowing formal decision theoretic approaches to be used to formal account for uncertainties in decision-making. Implementable approaches, methods, and guidelines will be presented for handling these topics.

Probabilistic uncertainty analysis is also called Bayesian statistical analysis, or simply Bayesian analysis, when Bayes rule is used to update estimated uncertainties with observed data. However, as will be described probabilistic uncertainty analysis is more general and can be applied without observational data. It is greatly enhanced and rendered more powerful with its systematic analysis of observational data to maximally reduce uncertainties. As part of the methods presented, the treatment of modeling uncertainties due to alternative, applicable models is described. The use of observed data in comparing with different model predictions is also described along with the construction of a consensus model prediction that maximizes the probability of a correct prediction. Numerous example applications involving MMOD risk predictions will be presented using spreadsheets and the software code WINBUGS which is freely available and which has a NASA sponsored user manual (1). The techniques and software algorithms though focusing on MMOD risk prediction examples are generally applicable. Numerous references are also provided giving details of the methods and techniques described.

Definition of Probabilistic Uncertainty Analysis

Probabilistic uncertainty analysis is the analysis of uncertainties by characterizing the uncertainty using a probability distribution. The probability distribution characterizing an uncertainty is also called an uncertainty distribution. Uncertainty distributions are most often applied to characterize possible estimated numerical values for a quantity. However, an uncertainty distribution can also be applied to characterize possible supports for alternative models which may be applicable. The uncertainty distribution can be continuous covering a continuous range of values or can be discrete covering specific

discrete values. Preciseness in the description of an uncertainty distribution is generally not necessary. Techniques are given for constructing an uncertainty distribution based on succinctly characterizing the uncertainty.

Characterizing uncertainties with probability distributions allows standard probability techniques to be used in propagating uncertainties in a model. The use of uncertainty distributions allows parameter uncertainties as well as model and assumption uncertainties to be systematically treated. Subjective uncertainties as well as data uncertainties are treated in a systematic manner. Equally important, observational data, even sparse and imprecise data, can be used to update and revise an estimated value and its associated uncertainty distribution. By updating with data, the estimate and its uncertainty are revised to be consistent with experience and are continually improved as experience is gained. Observational data can also be used to update and revise the support for alternative models thereby increasing the support for a model that best predicts the observed data. The examples that are described demonstrate the power of using observational data to update MMOD risk estimates and alternative model supports including observational data consisting of observations of zero MMOD occurrences.

Probabilistic uncertainty analysis is sometimes termed Bayesian analysis since Bayes rule is used to formally update and revise an estimated value and associated uncertainty with observational data. Because of its versatility, probabilistic uncertainty analysis is used in a wide variety of fields including risk and reliability analysis, forensic analysis, weather forecasting, legal evidence analysis, and information fusion analysis (2,3,4). Specific approaches and recommendations are given in this document for efficiently characterizing an uncertainty distribution associated with an estimated value. The examples focus on characterizing the uncertainty distribution associated with an MMOD risk estimate. Specific approaches are also given for statistically updating and revising an MMOD risk estimate and associated uncertainty distribution with observed data. Even though focusing on MMOD risk estimate applications, the approaches are generally applicable.

Steps Involved in Characterizing and Quantifying Uncertainties Probabilistically

There are basic steps involved in characterizing and quantifying parameter uncertainties and modeling uncertainties in a probabilistic manner. Parameter uncertainties are uncertainties in the value assigned to a value due to lack of knowledge of the precise value of the parameter. Modeling uncertainties are uncertainties in the model used to determine a result due to uncertainties in the relations and assumptions used.

Basic steps involved in characterizing and quantifying parameter uncertainties in a probabilistic manner involve:

1. Identifying the data sources of uncertainty
2. Estimating a central value , or best estimate value, for the parameter
3. Estimating a plausible range for the parameter value
4. Selecting a probability distribution, or uncertainty distribution, to describe the probabilities of possible values of the parameter
5. Updating and revising the uncertainties and uncertainty distribution of the parameter with new information and data:
 - a. Statistical updating and revision with observed data

- b. Updating the parameter characterization based on a new parameter definition

Methods and techniques are described in this document that can assist in carrying out these steps.

Basic steps involved in characterizing and quantifying modeling uncertainties in a probabilistic manner involve:

1. Identifying the contributors to the modeling uncertainty
2. Characterizing the effects of the contributors on the result
3. Postulating plausible different sizes of the contributor effects on the result
4. Estimating a central value, or best estimate value, for the result based on the plausible different sizes of the contributor effects
5. Estimating a plausible range for the result based on different plausible sizes of the contributor effects
6. Selecting a probability distribution, or uncertainty distribution, to describe the possible values of the result
7. Updating and revising the uncertainties and uncertainty distribution of the result with new information and data:
 - a. Statistical updating and revision with observed data
 - b. Model updating based on new models and assumptions

Methods and techniques are also described that can assist these steps.

Differentiating the Quantity to be Estimated Versus Its Estimated Value

In carrying out a probabilistic uncertainty analysis, it is first of all important to differentiate the quantity to be estimated from a particular possible value for the quantity. In probability theory, the quantity to be estimated is termed a random variable which can assume any of the possible values with associated probabilities. For example, the quantity to be estimated may be the probability of failure which can have particular numerical values. Or the quantity to be estimated may be an occurrence rate which can have particular occurrence rate values. The uncertainty distribution gives the probabilities for the possible values that the quantity can have. In terms of denoting the quantity as a random variable the uncertainty distribution gives the probabilities for the possible values that the random variable can have. When it is necessary, the quantity to be estimated will be differentiated from a particular value the quantity may have. Otherwise, whether it is the quantity or a particular value will be evident from the context of use.

Characterizations of an Uncertainty Distribution

As was indicated, the uncertainty distribution may be discrete or continuous depending upon whether the possible values of the estimated quantity are treated as being discrete or continuous. A discrete distribution is used when the estimated quantity has discrete possible values. A continuous distribution is used when the quantity is treated as having a continuous range of possible values. The range may be finite or infinite. Values for estimated quantities such as an estimated probability, an estimated occurrence rate, or estimated coefficients of a regression relationship are generally treated as being continuous. Discrete probabilities are used for a given number of failures that may occur in a mission or a given number of MMOD impacts that may be experienced. Discrete probabilities are also used to

characterize the relative supports for alternative models that may be used to determine an estimate or prediction.

Whether it is discrete or continuous, the uncertainty distribution is technically described by its probability density function (pdf) and cumulative distribution function (cdf), or equivalently the complementary cumulative distribution function (ccdf). . (For a discrete distribution, the term “probability mass function” is generally used instead of “probability density function”.) The probability density function (pdf) gives the probability that the quantity has a specific value. For a continuous distribution, the more formal definition for the pdf is the probability that the quantity has a specific value per unit value. This is due to the property that the pdf is multiplied by an infinitesimally small interval to give the probability for the value lying in the interval about the value. The “per unit value” will be omitted but be understood.

The cumulative distribution function (cdf) gives the probability that the quantity has a value less than or equal to a given value. The cumulative distribution is the integral of the pdf over all values less than or equal to a given value. (For a discrete distribution, the cumulative distribution function is the sum of the probability mass function (pmf) over all values less than or equal to a given value.) The complementary cumulative distribution function (ccdf) is one minus the cdf and is the probability that the quantity has a value greater than a given value. The cdf and ccdf are obtainable from the pdf and vice versa. In the following, the specific representation of the distribution, whether pdf, cdf, or ccdf, will be used when it is necessary. Otherwise the term “distribution” will be used to denote any of these equivalent representations.

Physical Uncertainty Versus Knowledge Uncertainty

There are two basic types of uncertainty-physical uncertainty and knowledge uncertainty. Physical uncertainty is uncertainty or variation due to a physical cause. Examples are weather variation versus time and location, material strength variation due to variation in the material processing, variation in the time of failure of a piece of equipment, and variation in the number of micrometeoroid hits experienced in a spaceflight. Physical uncertainty is also called aleatory uncertainty or irreducible uncertainty in that the physical uncertainty remains even if more knowledge is gained about the process.

Knowledge uncertainty is uncertainty due to lack of knowledge of the behavior, cause, or process. Knowledge uncertainty of an estimated quantity is uncertainty due to uncertainty in the model or in the parameter values used by a model. Examples of parameter uncertainties include uncertainty in the estimated failure rate of a piece of equipment and uncertainty in the estimated occurrence rate for the number of occurrences of a given event. Examples of modeling uncertainty include uncertainty in the model used to estimate the failure probability of a piece of equipment and uncertainty in the model used to estimate the probability of an MMOD hit on a spacecraft. Knowledge uncertainty is also called epistemic uncertainty or reducible uncertainty since the uncertainty can be reduced with more knowledge, e.g. more accurate modeling and greater data.

The same probabilistic uncertainty analysis approaches are used for both physical uncertainties and knowledge uncertainties. This is an advantage of quantifying uncertainties using probabilistic

uncertainty analysis. For certain applications, it is useful to separate the physical uncertainty contribution and the knowledge uncertainty contribution. A subsequent section describes approaches for doing this.

Specific Characteristics of an Uncertainty Distribution

The specific characteristics of an uncertainty distribution include the median, mean, variance, standard deviation, and specific quantiles or bounding values. These characteristics are standardly used in describing a specific uncertainty distribution and are defined as (5):

Median: the midpoint value of the uncertainty distribution. There is equal probability (50%) that the value is below or above the median value.

Mean: the weighted average of the possible values. Each possible value is weighted by its probability and then summed (integrated) to give the mean value.

Variance: a squared measure of the spread of the uncertainty distribution. The variance is the weighted average of the square of the distance of each possible value from the mean value.

Standard Deviation: a measure of the spread of the uncertainty distribution. The standard deviation is the square root of the variance.

Quantile or Bound Value: the value for which there is a given probability that the quantity is less than the value. Common bound values are 5% and 95% bound values. The median is a 50% bound.

To represent the characteristics in terms of formulas involving the uncertainty distribution, let

$$f(x) = \text{the probability density function (pdf) at a value } x \quad (1)$$

and

$$F(x) = \text{the cumulative distribution function (cdf) at a value } x. \quad (2)$$

Also let

$$M = \text{the median of the uncertainty distribution} \quad (3)$$

$$\mu = \text{the mean of the uncertainty distribution} \quad (4)$$

$$V = \text{the variance of the uncertainty distribution} \quad (5)$$

$$s = \text{the standard deviation of the uncertainty distribution} \quad (6)$$

and

$$x_p = \text{a given quantile or bound of the uncertainty distribution.} \quad (7)$$

Then the characteristics are given by;

$$F(M) = 0.50 \quad (8)$$

$$\mu = \int_R xf(x)dx \quad (9)$$

$$V = \int_R (x - \bar{x})^2 f(x) dx \quad (10)$$

$$s = \sqrt{V} \quad (11)$$

and

$$F(x_p) = p \quad (12)$$

where R is the range of values and p is the quantile or bound probability, e.g. 0.95.

Estimating an Uncertainty Distribution for a Basic Quantity

In practice, an uncertainty distribution for an uncertain quantity, such as an input parameter to a model, can be defined by three characteristics :

1. The estimated central value for the quantity
2. An estimated upper bound value, e.g. a 95% upper bound, for the quantity
3. The assigned distribution quantifying the probabilities of possible values of the quantity.

The Estimated Central Value

The central value for the quantity is generally associated with the median value or mean value of the uncertainty distribution. As described previously, the median value is the unbiased midpoint value of the possible values for the quantity in that there is equal probability (50%) of the quantity having a value lying below or above the median value. The mean value is a weighted average of the possible values weighted by their probabilities. Because the mean is a weighted average it is difficult to estimate without supporting calculation. The mean can also be biased in being a value higher or lower than the median. The central value is thus generally more accurately estimated as being the median value of the distribution. In estimating the central value as a median value, care should be taken that the value is not biased in being an overestimate or underestimate of the quantity. A moderate lack of precision in the central value, e.g. the central value being between the 40% and 60% bounds, is generally not an issue if the imprecision is included in estimating the uncertainty spread.

The Estimated Upper Bound Value

The estimated upper bound value has an associated coverage probability which gives the probability of the quantity being less than the upper bound value. The coverage probability is equivalently the confidence level that the quantity is less than the upper bound value. Instead of an estimated upper bound value an estimated error factor can be assigned. The estimated error factor is defined such that the upper bound value is obtained by multiplying the median value by the error factor. The upper bound value, or equivalently the error factor, is commonly defined so as to have an approximate 95% coverage probability. The precise value of the coverage probability is generally not an issue in characterizing the general size of uncertainty, e.g. being between 85% and 99%. It is, however, important that the coverage probability is not significantly overestimated. Past studies have shown that there is often

overconfidence in the estimate of a 95% upper bound or associated error factor and is more likely to be nearer to a 70% or 75% upper bound and associated error factor. If there is little basis for the upper bound being a 95% upper bound then a 75% value should be used for the coverage probability. Sensitivity studies can also be performed.

In certain cases it may be more convenient to estimate a lower bound value and associated error factor where the lower bound is obtained by dividing the median value by the error factor. The lower bound or associated error factor then have a specified exceedance probability which is the probability that the quantity has a value greater than the lower bound value. Also in certain cases it may be more convenient to estimate both a lower bound value and an upper bound value instead of estimating a central value and upper bound. The upper bound will then have a specified coverage probability and the lower bound will have a specified exceedance probability. Similar considerations as above apply in defining these alternative bounds and associated coverage probabilities.

The Assigned Distribution Family

The assigned distribution family is generally selected from one of the commonly used distributions which have enough flexibility to cover most uncertainty behaviors. The parameters of these distributions are also efficiently determined using the estimated central value and upper bound value, or their equivalents. The commonly used distributions, which are recommended for most problems, are:

1. The beta distribution
2. The uniform distribution
3. The normal distribution
4. The gamma distribution
5. The lognormal distribution.

All these distributions are standardly expressible in terms of two standard parameters which are given in the following descriptions. These two parameters can be determined from a specified central value and a specified upper bound value or their equivalents. Spreadsheets and software are available to do this determination. One spreadsheet program that was used for this document was developed by MD Anderson Center and is termed Parameter Solver and is available on the Internet (6). The formulas for the pdfs of these distributions in terms of their standard parameters are given in the following (7).

Beta Distribution

The beta distribution is used to describe a quantity which has a definite absolute minimum lower bound and definite absolute maximum upper bound. These bounds are generally (0, 1) but can be any values. The beta distribution is generally used to describe the uncertainty in an estimated probability or estimated reliability, among other applications. Equation (13) gives the standard formula for the beta probability density function (pdf) in terms of the two parameters (α , β):

$$f(x) = \frac{x^{\alpha-1} (1-x)^{\beta-1}}{B(\alpha, \beta)} \quad (13)$$

In the above formula $B(\alpha, \beta)$ is the normalizing beta factor. Figure 1 illustrates the different types of relative behaviors that can be described including continually increasing, unimodal, and u-shaped. In addition to the behaviors shown, a special case of the beta distribution is a uniform distribution which is described in the next section. Having specified a central value and upper bound value, the corresponding uncertainty distribution can be determined from the associated (α, β) parameters.

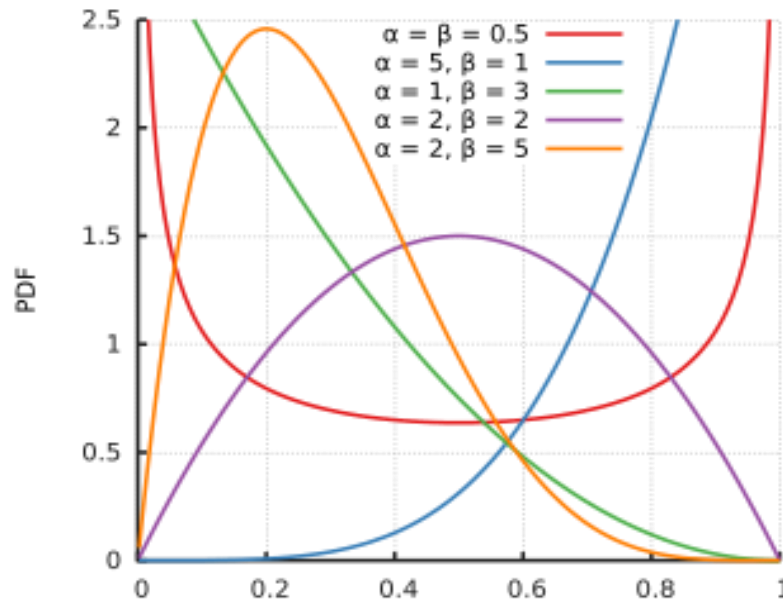


Figure 1. Different Shapes of the Beta Probability Density Function (pdf)

Uniform Distribution

The uniform distribution is a special case of the beta distribution ($\alpha=1, \beta=1$) and characterizes all possible values in a defined range as being equally likely. The shape of the uniform probability distribution function (pdf) is thus flat across the range; in Figure 1 it is a horizontal line at 1. The uniform distribution either on a linear scale or log scale is occasionally used as a non-informative uncertainty distribution portraying minimal information about the uncertainty. However, if there are no natural defined finite range application can be limited by requiring one to be defined. Sensitivity studies can be carried out to determine the sensitivity of the results to the bounds by varying the bounds. The gamma or lognormal which do not have this definite bound limitation can also be used to characterize a non-informative uncertainty by specifying a large error factor, e.g. a factor of 10, which is associated with the central value. These distributions will be subsequently described.

Normal Distribution

Equation (14) gives the standard equation for the normal probability density function (pdf) in terms of the mean and standard deviation (μ, σ):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (14)$$

Figure 2 illustrates the different shapes the normal probability distribution (pdf) can have. Because the range of the normal distribution generally covers both negative and positive values, it is an applicable uncertainty distribution when the possible values can be both negative and positive. When the possible values can only be positive, such as for an estimated probability value or occurrence rate value, then the range needs to be truncated. When the uncertainty range is such that negative values have a low likelihood then the normal distribution can be used as an approximate uncertainty distribution without truncation. The gamma distribution or lognormal distribution is more easily handled as uncertainty distributions for positive quantities since they cover only positive values and do not need to have artificial truncation constraints imposed.

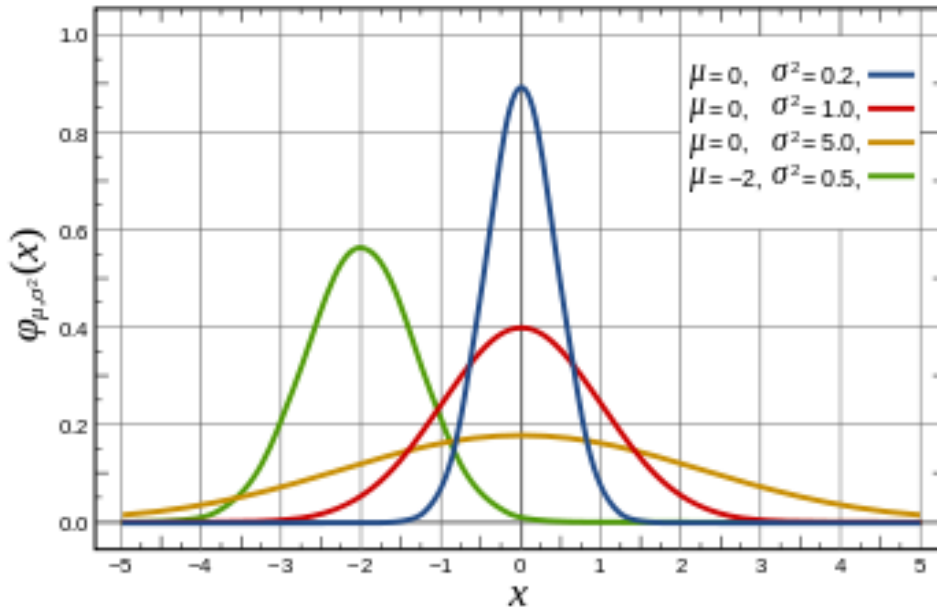


Figure 2. Different Shapes for the Normal Probability Density Function (pdf)

Gamma Distribution

Equation (15) gives a standard equation for the gamma probability density function (pdf)

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x). \quad (15)$$

In the above formula $\Gamma(\alpha)$ is the normalizing gamma function. Figure 3 illustrates the different shapes the gamma distribution (pdf) can have. The gamma distribution covers a positive range with zero as a lower limit and no defined upper limit. It therefore is useful as an uncertainty distribution for positive quantities such as a failure rate estimate or occurrence rate estimate. A specified upper bound, e.g. a

95% upper bound, controls the upper values. The gamma distribution has the useful feature that its two parameters can be explicitly updated from observation data of occurrences or non-occurrences of events. This is a particularly useful feature in easily updating an occurrence rate estimate or failure rate estimate along with its uncertainty distribution as observational data is collected. For larger error factors, e.g. a 95% error factor of 10, the gamma distribution can give relatively large probabilities to values near zero as compared to the lognormal. However, as will be shown, there is generally little difference between the gamma and lognormal distribution values greater than the median if both are pegged to have the same median value and same 95% error factor.

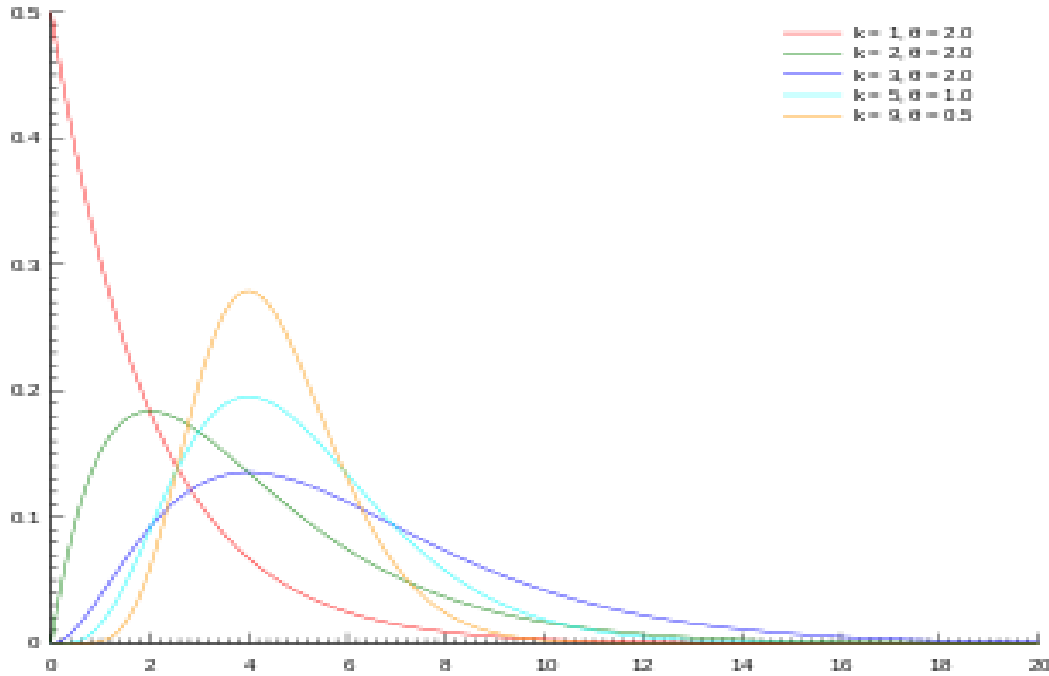


Figure 3. Different Shapes for the Gamma Probability Density Function (pdf)

Lognormal Distribution

Equation (16) gives the standard equation for the lognormal probability density function (pdf):

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right) \quad (16)$$

Where μ and σ are the mean and standard deviation of the lognormal distribution.

Figure 4 illustrates the different shapes the lognormal distribution (pdf) can have. The lognormal covers a positive range for the estimated quantity from zero to no defined upper limit. The lognormal shapes are generally similar to those for the gamma where the gamma has higher probabilities at values near

zero for the same median and 95% error factor. The lognormal distribution is the distribution commonly used for the uncertainty distribution in risk and reliability estimates for a failure rate estimate or occurrence rate estimate. The lognormal has the feature that it transforms to a normal distribution for the log of the quantity. The lognormal distribution is thus a natural distribution to use when the log of the quantity is the natural scale. The lognormal has the property that the median value times the error factor and divided by the 95% error factor give symmetric upper 95% and lower 5% bounds about the median. . As for the gamma distribution, when the 95% error factor is large, e.g. 10, the lognormal distribution is fairly flat over the range representing a non-informative uncertainty.

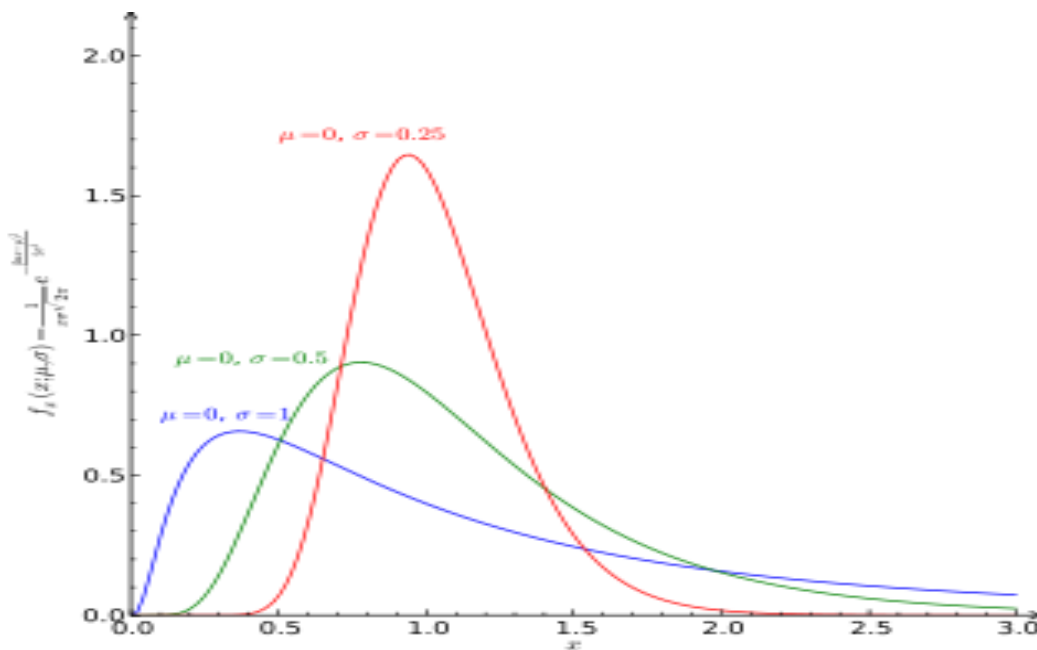


Figure 4. Different Shapes for the Lognormal Probability Density Function (pdf)

Propagating Input Uncertainties in a Model

The first step in propagating uncertainties in a given model involves estimating an uncertainty distribution to each of the input quantities in the model where the input quantities are generally termed the input parameters of the model. For each input parameter which has associated uncertainty, an uncertainty distribution is estimated using the previously described steps, or using equivalent steps, for characterizing a specific uncertainty distribution. Once uncertainty distributions are estimated for each of the input parameters, the uncertainty distributions are then propagated through the model to obtain the uncertainty distribution of the evaluated result.

The mechanics of propagating uncertainty distributions is straightforward using available software. Monte Carlo simulation is standardly performed for other than simple explicit equations relating the

final result value to input values. Monte Carlo simulation can be performed directly on the model or can be performed using a response surface representation of the model. When Monte Carlo simulation is directly performed on the model then the Monte Carlo simulation software is used as a front end input selector to a model. A set of input values is randomly selected (sampled) according to their uncertainty distributions and are input to the model to determine a particular output result value. This is repeated a number of times to obtain a random sampled set of result values from which the uncertainty distribution characteristics are estimated, e.g., the median, mean, and bounds. For more efficient simulation, the sampled input values can be selected using a design of experiments type to direct the sampling, such as using Latin Hypercube sampling. For a complex model, a response surface is alternatively fit. Direct Monte Carlo simulation is then performed on the response surface representation of the model output. Response surface model simulation requires the extra step of determining the response surface but can save considerable computer time if the model calculations are complex and are time consuming. References 8 and 9 give more details on the methods and tools that are used for propagating uncertainty distributions.

Common Ways of Handling Dependencies Among Uncertain Quantities

When values are estimated for multiple input parameters, or input quantities, that have uncertainties then the possibility of dependencies among the estimated values need to be considered. Dependency among estimated values occurs if the value estimated for one parameter depends on the value estimated for other parameters. An example that occurs in system reliability modeling is using one estimated failure rate for all similar components, e.g. for all check valves of a given type. Another example is the estimation of structural failure probabilities at different locations of a spacecraft that are made of the same material and that experience similar stresses. As these examples indicate there are two common cases when dependencies among uncertain parameters or uncertain quantities need to be handled. The first case occurs when the same value is used for estimates of multiple quantities. The second case occurs when multiple the quantities are functionally related to a smaller set of causal or influencing quantities.

As indicated, the first case occurs when multiple check values are given the same failure rate uncertainty distribution. The apparent multiple failure rate estimates and multiple failure rate uncertainty distributions are thus in essence only one estimate with one uncertainty distribution. In carrying out a Monte Carlo simulation, when a failure rate is sampled from the uncertainty distribution for a check valve then all the check values in the set are given the same failure rate value. This situation is commonly called complete coupling of the estimates. This case differs from the situation where the failure rates are independent but have the same uncertainty distribution. In Monte Carlo simulation, the failure rates are independently sampled from the same distribution. The check valve failure rate estimates are independent, for example, when the uncertainty in a failure rate value is dominated by the physical variability in the component failure rate due to manufacturing variability or local environmental variability. If it is determined that this variability randomly and independently occurs among the components then the estimated values and uncertainty distributions are treated as being unique and uncoupled. Otherwise, if the physical variabilities are systemic and coupled in nature, e.g., due to same overall environment, then a single uncertainty distribution needs to be used..

In the second common case multiple quantities are functionally related to common underlying causal variables. The relationship and dependency can be expressed in terms of a probability relationship or in terms of a regression relationship. In a probability relationship, the probability of the multiple quantities having specific values is expressed as the product of probability of the common causal parameters having specific values times the probability that the multiple quantities have their specific values given the values of the causal parameters. An example is the estimation of the common cause failure of multiple susceptible equipment. If the common failure cause or systemic effect exists then with a given probability all the susceptible equipment fail. This conditional common cause failure probability is estimated as being one in the extreme case. If the common cause does not exist then the multiple equipment independently fail. The overall probability of failure is then the probability of the common cause existing or no existing times the respective conditional failure probabilities of the equipment given the respective case. This second common case also applies when multiple quantities are related to causal variables through explicit regression equation relationships a general approach and associated tools for handling these types of dependency is a Bayesian net which links causal parent variables to a causal variable through conditional probability tables or formulas (10, 11). When uncertainties are assessed then uncertainty distributions are assigned to the causal variables (causal parameters) and are propagated through the relationships.

Considerations Involved in Quantifying the Uncertainty in an MMOD Risk Estimate

Before proceeding in describing the handling of modeling uncertainties due to different, alternative models, it is useful to present an example of quantifying the uncertainty in an MMOD risk estimate. This is a simple example but it illustrates some of the considerations that are involved. The MMOD risk estimate is defined here as the probability of failure of a spacecraft in a mission due to a failure-causing MMOD impact. The probability of no MMOD failure one minus the probability of failure and is the success probability and reliability.

The probability of no MMOD failure impact to a spacecraft can be concisely represented as a product of probabilities where each probability is the probability of no MMOD failure at a given location on the spacecraft in a given time interval. Modeling the probability of MMOD impact as a Poisson probability, the probability of no MMOD failure of the spacecraft in mission can be represented as

$$R = \prod_{i,j} \exp(-\Lambda(x_i^p, t_j)) \quad (17)$$

where

$$R = \text{the probability of no spacecraft failure due to an MMOD impact} \quad (18)$$

$$\Lambda(x_i^p, t_j) = \text{the Poisson occurrence rate for a failure causing MMOD impact on an area} \quad (19)$$

centered at x_i^p and during a time interval centered at time t_j

and

$$\prod_{i,j} = \text{the product over all locations } i \text{ and times } j \text{ during the mission} \quad (20)$$

Because of the specific trajectory of the spacecraft, the location x_i^p is a function of the time t_j . The number of product terms in Equation (17) is can be very large and depends on the fineness of the grids used for the discrete location areas and time intervals. Each Poisson failure occurrence rate $\Lambda(x_i^p, t_i)$ is in turn a product of the MMOD impact occurrence rate and the conditional probability that the MMOD impact penetrates and causes failure. The probability that the MMOD impact causes failure depends on the impacted material's characterization and in detailed modeling is evaluated using finite

In constructing an uncertainty distribution for the probability of no failure R , the detailed approach is to construct the uncertainty distributions for the set of MMOD Poisson failure occurrence rates $\Lambda(x_i^p, t_i)$ based on the relationships and input parameters determining their values. The estimated values for the occurrence rates $\Lambda(x_i^p, t_i)$ are not independent since they depend on common parameters and dependent MMOD impacts at nearby locations and time intervals. Uncertainty distributions are assigned to the independent parameters and propagated to obtain the uncertainty distributions for the set of occurrence rates at the grid of locations and time intervals. This is performed using Monte Carlo simulation. Using Equation (17) the sampled occurrence rates $\Lambda(x_i^p, t_i)$ determine a sample value for the probability of no failure R . The sampling is carried out for a number of trials obtaining a set of values for R (and also for the failure probability $1 - R$). An empirical probability distribution (can then be constructed from the set of sampled values.

The above detailed approach for determining the uncertainty distribution for R works for simpler models but becomes less feasible for more complex problems because of the computer simulation time needed and the difficulty in constructing the uncertainty distributions for the input parameters that determine the values for the set of Poisson occurrence rates $\Lambda(x_i^p, t_i)$. Because of the difficulty in constructing the uncertainty distribution with the associated uncertainties in the assumptions that need to be made, it can be more effective to use a simpler, approximate model for the probability of no failure in order to estimate the associated uncertainty distribution. Since the uncertainty distribution is used to assess the general size of uncertainty and its dominant contributors, preciseness in detailed shape and values is generally not a priority. Also, as will be demonstrated in a subsequent section, observational data, even sparse data, can be used to update and revise the initial uncertainty approximation to be aligned with actual data. Consequently, initial preciseness in the uncertainty will be overridden with the updating and revision from observational data. To circumvent the detailed evaluations which have large uncertainties in themselves, subjective, informed judgments can be used to estimate the uncertainty distribution.

In constructing a more workable uncertainty formulation, it is useful to re-express the Equation (17) for the probability of no MMOD failure as

$$R = \exp\left(-\sum_{i,j} \Lambda(x_i^p, t_j)\right) \quad (21)$$

where the summation is over the locations and times covered in the detailed modeling. For a more workable evaluation, the summation can then be approximated by a more limited sum of only those

highest failure rates $\Lambda(\mathcal{P}_i, t_i)$ at critical locations \mathcal{P}_i and critical times t_i . The uncertainty of the summation of failure rates in Equation (21) can be fit with a lognormal distribution having the mean and variance determined from the sum of means and variances of the individual failure rate uncertainty distributions. Since each individual failure rate is a product of the MMOD impact occurrence rate and the probability of penetration failure given an impact then using lognormal uncertainty distributions for each factor the product also has a lognormal uncertainty distribution. The characteristics of these distributions can in turn be estimated using simpler response surfaces for the occurrence rates.

Handling of Alternative Models and Associated Modeling Uncertainty

In particular situations alternative models may exist for calculating a result such as a predicted risk value. The alternative models may not have the same support but there is enough support in each model to consider it to some degree. The principles of probabilistic uncertainty analysis provide a systematic method for quantifying the modeling uncertainty and for combining the alternative model results in an optimal fashion to obtain a consensus result and associated uncertainty. By optimally combining the individual model results to obtain a consensus result the probability is maximized that true value is near the consensus result and that it is contained in the associated bounds (12, 13).

The steps involved are in identifying and quantifying model uncertainty involve:

1. Identifying the alternative models
2. Determining the results from the alternative models
3. Determining the uncertainty distribution for each model result
4. Optimally combining the individual models results and their uncertainties

As an example, a current MMOD risk model contains an additional flux contribution from stainless steel debris. The stainless steel flux contribution is based on one set of assumptions and relationships. The set of assumptions and relationships has high uncertainty because of the lack of evidence and validation for the additional flux contribution. Other sizes of the stainless steel flux contribution can be obtained by different assumptions and relationships. Each set of assumptions and relationships provides a different MMOD risk model. As one limiting model, the additional flux contribution can be excluded giving an MMOD model with improved covering all other flux contributions that have been modeled in the past. By not considering a sampling of these other models and their results, it is not clear that the one set of assumptions and relations provides an upper bound since exclusion of the contribution always provides a lower risk estimate which is significantly lower in certain cases.

In general, by explicitly considering alternative model results, a range on the possible results is obtained due to modeling uncertainty. The range now includes not only the within-model uncertainty but also the across-model uncertainty. As in any statistical sample, the alternative models do not have to be exhaustive but instead be representative of the different possibilities. As will be shown, differences in the supports for the alternative model results can be factored in constructing the consensus result. Observed data can also be used to update and revise the supports by increasing those that are consistent with the data and decreasing those that are not. We consider two alternative models that provide estimates of the same quantity such as the risk to a spacecraft for a given mission. The generalization to multiple models follows in a straightforward manner.

Let

$$P(M_1) = \text{the relative support for Model 1} \quad (22)$$

$$P(M_2) = \text{the relative support for Model 2} \quad (23)$$

where

$$P(M_1) + P(M_2) = 1. \quad (24)$$

Formally, the relative supports $P(M_1)$ and $P(M_2)$ are the probabilities that Model 1 or Model 2, respectively is the applicable model and are the relative confidences in the particular models. The relative supports are assigned based on experience, verification and validation of the models. Criteria can be used in assisting in the determination of the relative supports such as identified in NASA Standard 7009 (14) which gives rating criteria for the support of a model. The ratings in 7009 for a model can be summed and can then be normalized across the two models to give the relative supports. Equal supports can be used when there are countering arguments for each model. The relative supports do not need to be precise especially if they are updated and are revised based on observed data as will be described. The formulation also allows sensitivity studies to be simply performed.

To continue the formulation let

$$f_1(y) = \text{the probability that the estimated quantity has value } y \text{ according to Model 1} \quad (25)$$

and

$$f_2(y) = \text{the probability that the estimated quantity has value } y \text{ according to Model 2.} \quad (26)$$

The distributions $f_1(y)$ and $f_2(y)$ are equivalently the uncertainty distributions (pdfs) for the quantity having value y determined from Models 1 and 2, respectively. The uncertainty distributions are equivalently expressed as probabilities here to emphasize their giving the probability of a given value for the estimated quantity of interest.

Also let

$$f(y) = \text{the consensus probability that the estimated quantity has value } y \quad (27)$$

by combining the evaluations from both Models 1 and 2.

The consensus probability $f(y)$ is more formally the consensus pdf for the quantity having value y and is the consensus uncertainty distribution. Then the consensus probability distribution $f(y)$ that maximizes the probability of an accurate estimation of the value of the quantity is given by

$$f(y) = P(M_1)f_1(y) + P(M_2)f_2(y). \quad (28)$$

The consensus probability distribution $f(y)$ is thus the weighted average of the individual model probability distributions $f_1(y)$ and $f_2(y)$. In formal probability theory $f(y)$ is termed the mixed model probability distribution since the model distributions are mixed to obtain the consensus distribution. It is important to note that the consensus estimation distribution $f(y)$ is the weighted average of the individual model probability distributions $f_1(y)$ and $f_2(y)$. Having determined the consensus distribution using Equation (28) it then can be used to obtain consensus estimates such as the consensus mean, the consensus median, and consensus upper and lower bounds.

The individual model distributions $f_1(y)$ and $f_2(y)$ are obtained by propagating the input uncertainty distributions of the parameters to the individual model as previously described. As also described the individual distributions can also be obtained by directly setting bounds on the individual results and fitting the uncertainty distribution. Each model distribution considers only the uncertainties from the contributions considered within the model. Contributions in the other model are not considered. The consideration of the different model contributions and associated uncertainties are taken into account in the combination (mixing) of distributions given by Equation (28).

If one model, for example Model 1, has total support then $P(M_1) = 1$ and $P(M_2) = 0$. The consensus distribution is then simply the distribution of Model 1. By allowing for the possibility that an alternative model estimate is the more accurate one the consensus estimate maximizes the probability of the estimate being correct. As indicated, the relative supports will be updated and be revised with observed data resulting in the support being increased for the model which is most consistent with the data.

Various techniques can be applied in using Equation (28) to determine the consensus uncertainty distribution from the individual model uncertainty distributions. If Monte Carlo simulation is used then a particular model is first selected according to the relative supports (probabilities) $P(M_1)$ and $P(M_2)$. Given the particular model selected a value for the estimated value y is sampled from the particular model uncertainty distribution $f_1(y)$ or $f_2(y)$. This is repeated a number of times to obtain the sampled values of y which are then used to estimate the consensus uncertainty distribution. Various simulation codes can be used to carry out these evaluations. If the model uncertainty distributions are ones of the standard distributions that have been described, or are fitted to one of these distributions, then the simulations can be efficiently carried out including using EXCEL®.

Equation (28) can also be directly used to obtain specific characteristics of the consensus distribution which can then be used to fit one of the standard distributions such as the lognormal. Applying probability procedures to Equation (28), the consensus estimated mean value is given by

$$\mu = P(M_1)\mu_1 + P(M_2)\mu_2 \quad (29)$$

where

$$\mu = \text{consensus mean value} \quad (30)$$

$$\mu_1 = \text{mean value from Model 1} \quad (31)$$

$$\mu_2 = \text{mean value from Model 2.} \quad (32)$$

The variance of the consensus estimate is the sum of two contributions

$$V = V_w + V_A \quad (33)$$

where

$$V = \text{the variance of the consensus estimate} \quad (34)$$

$$V_w = \text{the average of the within model variances} \quad (35)$$

$$V_A = \text{the variance across the model predictions} \quad (36)$$

The average of the within model variances V_w is given by

$$V_w = P(M_1)V_1 + P(M_2)V_2 \quad (37)$$

where

$$V_1 = \text{variance of } y \text{ from the uncertainty distribution of Model 1} \quad (38)$$

$$V_2 = \text{variance of } y \text{ from the uncertainty distribution of Model 2} \quad (39)$$

The within variance contribution is thus the weighted average of the individual model variances for the quantity. The across variance contribution V_A accounts for the difference in the models in terms of their different means from their uncertainty distributions:

$$V_A = P(M_1)(\mu_1 - \mu)^2 + P(M_2)(\mu_2 - \mu)^2. \quad (40)$$

Using these relationships, the determined consensus mean value μ and consensus variance V can then be used to fit an uncertainty distribution such as the lognormal. This fitted uncertainty distribution will not necessarily show detailed structure of the more precisely determined uncertainty distribution, which may be bimodal if the individual model uncertainty distributions are widely separated. However, it will accurately identify the central values of the mean and median and the overall spread of the consensus estimate which includes the different model estimates and their uncertainties.

Example of Handling Alternative Models in an MMOD Risk Estimate

It will be useful to present an example of alternative, candidate models that is relevant in the prediction of MMOD risk to a spacecraft. The example will also show how the consensus model equation, Equation (28), can be used to determine the consensus risk prediction. As was indicated, one can view two alternative, candidate models as being available to use to predict the risk of MMOD failure to a spacecraft in its mission. One MMOD model which we will denote as MMOD+ includes an orbital debris model (ORDEM3.0) which includes an orbital debris contribution from high density stainless steel debris in the size range of 1 to 3 mm. This contribution can produce high failure risk contributions in particular spacecraft missions but it has high uncertainty and has not been validated. The other model which we will denote as MMOD- includes all the orbital debris contributions in ORDEM3.0 except the high density stainless steel orbital debris contribution.

As was given by Equation (21) the probability of no MMOD failure can be expressed as

$$R = \exp\left(-\sum_{i,j} \Lambda(\mathcal{P}_i, t_j)\right). \quad (41)$$

For two alternative model estimates, the consensus probabilities (pdfs) given by Equation (28) also applies to with the cumulative distribution functions (cdfs) (or equivalently the complementary cumulative distribution function (ccdfs) substituted for the pdfs. Since the probability of no failure is a reliability function and hence a ccdf the consensus risk estimate using the two alternative models is

$$R = P(M_+) \exp\left(-\sum_{i,j} \Lambda_+(\mathcal{P}_i, t_j)\right) + P(M_-) \exp\left(-\sum_{i,j} \Lambda_-(\mathcal{P}_i, t_j)\right). \quad (42)$$

where the $P(M_+)$ ($P(M_-)$) denotes the relative support for the MMOD+ model (MMOD- model) and $\Lambda_+(\mathcal{P}_i, t_j)$ ($\Lambda_-(\mathcal{P}_i, t_j)$) denotes the MMOD failure rates for the MMOD+ model (MMOD-model). As was previously discussed, the uncertainty distributions from the alternative model estimates can then be combined using Equation (42) to obtain the uncertainty distribution in the consensus probability of no MMOD failure.

Determining Uncertainty Contributions to an Overall Uncertainty

In various problems it is useful to determine more detailed uncertainty contributions to the uncertainty in a result as characterized by its associated uncertainty distributions. Even though an input parameter has a large uncertainty it may have a small contribution to the overall result uncertainty because of its small role in determining the value of the result. Conversely, a quantity may have small uncertainty but have a large contribution because of its importance to the overall result.

As earlier indicated, one useful measure of the size of uncertainty is the variance (and equivalently the standard deviation) of the uncertainty distribution. The previous relationships involving the model variance contributions to the overall variance can be extended to determine the variance contributions of any variables to an overall variance. Before describing this extended approach, which can be termed the conditional variance approach, it is useful to describe three alternate approaches that can be used in

certain cases to determine uncertainty contributions. These are the linear variance propagation approach, the Taylor approximation approach, and the uncertainty removal approach.

The linear variance propagation approach and the Taylor expansion approach are closely related. The variance propagation approach can be used if a result y can be expressed as a linear contribution of the uncertain parameters or variables. A simple example which represents the basic approach is

$$y = c_1x_1 + c_2x_2 + c_3x_3 + \dots + c_nx_n \quad (43)$$

where there are n contributors and x_i are the uncertain parameters and c_i are the respective coefficients. Then using standard variance propagation techniques the variance of y can be expressed as linear sum of the variances of x_i times their coefficients squared plus twice the covariances of all combinations of the parameters x_i and x_j that have estimation dependency times the product of their coefficients (15). If all the variables are independent or have small dependency then the variance of y is simply the sum of the individual variance contributions times the square of their coefficients. Because of the limitation of linearity, the linear variance propagation approach is limited to a narrow set of problems. Also, when a response surface is fitted to the results by varying the parameter values then the linear propagation approach is applicable only when the response surface is linear in the parameters.

The Taylor approximation approach is more general in that it approximates the variance of y by a linear sum of the variances and covariances of the parameters by suitably determining the first order coefficients in a linear approximation of the variance y . For the basic formulation, let

$$V_y = \text{the variance of the result } y \quad (44)$$

$$V_i = \text{the variance of the parameter } x_i \quad (45)$$

$$\frac{\Delta y}{\Delta x_i} = \text{the ratio of the change in } y \text{ per change in } x_i. \quad (46)$$

When the result is an explicit function of the parameters then the ratio of the change in y per unit change in a parameter is replaced by the partial derivative. Otherwise, and more generally, a small change in a parameter is made and the resulting change in the result is determined. The change in the parameter is selected to be small enough that the change in the result is accurately approximately by a first order linear expansion. The initial parameter values are all set at their mean values in determining the ratios. Then using a first order Taylor expansion for y the variance of y is approximated by

$$V_y \cong \left(\frac{\Delta y}{\Delta x_1}\right)^2 V_1 + \left(\frac{\Delta y}{\Delta x_2}\right)^2 V_2 + \dots + \left(\frac{\Delta y}{\Delta x_n}\right)^2 V_n + \dots \quad (47)$$

The Taylor approximation approach is only accurate when the uncertainties in the parameters x_i are small enough that a linear approximation for y is as the parameters vary within the ranges, e.g. within their 95% bounds. This generally limits the standard deviations of x_i to be a small fraction of their means, e.g., less than 10% of the respective means. For larger uncertainties in the parameters, the Taylor approximation approach can give results which are significantly in error and in particular can significantly underestimate the variance in the result y . Higher order terms can be added to the Taylor expansion but these involve higher order covariances of the product of the parameters which generally are not simple to determine (16).

In the uncertainty removal approach, the overall variance of the result of interest is first determined with all its contributors. Then the uncertainty of a selected contributor parameter is reduced to zero with the parameter set to its mean value. The overall variance is again evaluated and the reduction in overall variance is a measure of the uncertainty contribution from the parameter. This is repeated for each of the selected parameters whose uncertainty contribution is to be determined. To account for interactions among the variables, the uncertainties of sets of parameters can be simultaneously reduced to zero and the overall variance again evaluated to determine the resulting reduction.

The uncertainty removal approach is thus a brute force method. It is feasible for simpler models and for focused evaluations of selected uncertainty contributions. The uncertainty removal approach can also be applied to a simpler response surface model that is used to approximate a more complex model. Otherwise, the uncertainty removal approach is not feasible for more complex models because of the time consuming evaluations.

The conditional variance approach is the most comprehensive approach for determining the variance contributions and importances of the parameters. The approach is described in detail in reference (17). Basically, the conditional mean of the result is evaluated for a fixed value of the variable x_i by allowing the other parameters to vary over their uncertainty distributions. The contributions of the other parameters are thus averaged out in determining the conditional mean for a fixed value of a parameter. The variance of the conditional mean over the uncertainty distribution of x_i then gives the variance contribution of x_i . This is repeated for all the variables of interest. The interaction contributions can also be determined if desired. This approach is theoretically shown to give all the individual parameter variance contributions and all the parameter interaction contributions. The approach can be practically applied using efficient software developed by the Joint Research Center of the European Commission which is publicly available (17).

Separating Physical and Knowledge Uncertainty Contributions

The conditional variance approach described in the previous sections can be used to separate the physical uncertainty contribution from the knowledge uncertainty contribution. This can be useful in determining whether to spend resources to gain additional knowledge versus spending resources to change the physical situation to reduce the physical uncertainty. As a basic example, consider the time to failure of a piece of equipment. The total uncertainty in the predicted time of failure is due to both

the physical uncertainty, i.e. the randomness, in the time of failure if the failure rate were known plus the uncertainty in knowledge of the failure rate. Using the conditional variance decomposition described in the previous alternative model section, the total variance of the time to failure can be expressed as

$$V(t) = E(V(t / \lambda)) + V(E(t / \lambda)) \quad (48)$$

where

$$V(t) = \text{the total variance of the time to failure} \quad (49)$$

$$V(t / \lambda) = \text{the variance of the time to failure for a given failure rate value } \lambda \quad (50)$$

$$E(V(t / \lambda)) = \text{the expectation (mean value) of } V(t / \lambda) \text{ over the uncertainty distribution of } \lambda \quad (51)$$

$$E(t / \lambda) = \text{the mean value of the time to failure for a given failure rate value } \lambda \quad (52)$$

$$V(E(t / \lambda)) = \text{the variance of } E(t / \lambda) \text{ over the uncertainty distribution of } \lambda \quad (53)$$

The failure rate value λ is in general any parameter characterizing the time to failure distribution such as the failure rate parameter of the exponential distribution or the shape parameter of the Weibull distribution.

The first variance contribution $E(V(t / \lambda))$ on the right hand side of Equation (48) is the physical uncertainty contribution. If the failure rate value were known exactly then the expectation would simply be at the given value and the variance $V(t / \lambda)$ would be the physical variance in time of failure for the known failure rate. When the failure rate value is not precisely known then this physical contribution is averaged over the possible values of the failure rate.

The second contribution $V(E(t / \lambda))$ in Equation(48) is the knowledge uncertainty contribution. If the failure rate were known exactly this contribution would be zero. The expected value of the time to failure would be at the known failure rate value and the variance of the expected value would consequently be zero. When there is uncertainty in the failure rate value then the variance in the expected value represents the knowledge uncertainty.

When the failure time follows an exponential distribution then using the above general relationship, specific expressions can be obtained for the physical and knowledge uncertainty contributions. Assuming a given failure rate value for the exponential time to failure

$$V(t / \lambda) = \frac{1}{\lambda^2} \quad (54)$$

and

$$E(t / \lambda) = \frac{1}{\lambda} . \quad (55)$$

The variance contributions then become

$$V(t) = E\left(\frac{1}{\lambda^2}\right) + V\left(\frac{1}{\lambda}\right) . \quad (54)$$

The first contribution on the right hand side of Equation(54) is the average physical uncertainty contribution over the uncertainty distribution for λ . The second contribution is the knowledge uncertainty contribution expressed as the variance of the mean time to failure over the uncertainty distribution for λ . . The specific sizes of the these contributions, depend on the uncertainty distribution assigned to λ . The second contribution measuring the size of knowledge uncertainty is particularly sensitive to the spread of the uncertainty distribution as measured by the variance.

Similar decompositions can be performed for other physical quantities which have both a physical variability, i.e., physical uncertainty, and a knowledge uncertainty. Examples include variability in the an estimated material strength and variability in an environmental property or characteristic The general decomposition equation, Equation (46), can also be straightforwardly extended to cover the physical and knowledge uncertainty contributions from multiple parameters used to model the time to failure distribution. For example, the two Weibull parameters can be substituted for λ in which case the conditional means and variances are over the two parameter uncertainty distribution.

Revising an Estimate and Its Uncertainty Using Observed Data

A strong property of probabilistic uncertainty analysis is the use of probability principles to statistically revise, or update, an estimate and its uncertainty with observed data. The effect of the statistical revision depends on both the size of uncertainty associated with the estimate and the amount of observed data and its precision. The important point of statistical revision is that the estimate is directly revised and aligned with the observed data using probability and statistical techniques.

The statistical revision and updating is carried out using the general conditional probability relationship between two events A and B

$$P(A / B)P(B) = P(B / A)P(A) \quad (55)$$

where

$$P(A / B) = \text{the conditional probability of } A \text{ occurring given the knowledge of the occurrence of } B \quad (56)$$

$$P(B) = \text{the probability of } B \text{ occurring} \quad (57)$$

$$P(B / A) = \text{the conditional probability of } B \text{ occurring given the knowledge of the occurrence of } A \quad (58)$$

$$P(A) = \text{the probability of } A \text{ occurring.} \quad (59)$$

When this general relationship is applied to revise an uncertain estimate with observed data then the relationship is termed Bayes rule and the updating of the estimate is termed Bayesian statistical analysis or simply Bayesian analysis (18,19).

To apply Bayes rule let

$$P(E / v) = \text{the probability of the observed event } E \text{ occurring given the estimated quantity has value } v \quad (60)$$

$$P(v) = \text{the probability that the quantity has value } v \quad (61)$$

$$P(v / E) = \text{the probability that the quantity has value } v \text{ given knowledge of the observed event } E \quad (62)$$

$$P(E) = \text{the probability of the event } E \text{ occurring.} \quad (63)$$

Then the general conditional relationship, or Bayes rule, becomes

$$P(E / v)P(v) = P(v / E)P(E) . \quad (63)$$

Rearranging we have

$$P(v / E) = \frac{P(E / v)P(v)}{P(E)} . \quad (64)$$

Finally using the relationship expressing the total probability in terms of the conditional probabilities, $P(E)$ can be expressed as

$$P(E) = \int_{v'} P(E / v') \{P(v') dv' \quad (65)$$

Equation (65) says that the total probability $P(E)$ is the integral over the possible quantity values of the conditional probability of E for a given value of v' times the probability of the quantity having this value. The integral in Equation (65) is over all possible values v' . For discrete possible values the integral in Equation (65) is replaced by a sum over the possible values. Equation (64) can then be expressed as

$$P(v/E) = \frac{P(E/v)P(v)}{\int_{v'} P(E/v')P(v')dv'} \quad (66)$$

Equation (66) gives the general Bayes rule for updating an estimate and its uncertainty using observed data. Equation (66) is also commonly expressed

$$f(v/E) = \frac{L(E;v)f(v)}{\int_{v'} L(E;v')f(v')dv'} \quad (67)$$

where

$$L(E;v) = \text{the likelihood for the observed data } E \quad (68)$$

$$f(v) = \text{the prior distribution for the parameter} \quad (69)$$

$$f(v/E) = \text{the posterior distribution for the parameter given the data } E. \quad (70)$$

The prior distribution $f(v)$ is the prior uncertainty distribution for the parameter and the posterior distribution $f(v/E)$ is the posterior uncertainty distribution for the parameter. The observed data, or observed event, can be of any type. Table 1 gives examples of the observed event and likelihood function with associated parameter or parameters. The examples given in later sections on updating MMOD risk estimates will show how Bayes rule is practically applied.

Table 1. Examples of Applications of Bayes Rule to Statistically Revise Parameter Estimates

EVENT or DATA	LIKELIHOOD FUNCTION	PARAMETER(S)	PRIOR DISTRIBUTION
Component failures occurring at given times	Probability of failures occurring at given times using the exponential failure time distribution	The failure rate parameter in the exponential	Uncertainty distribution in the exponential failure rate
Number of occurrences of an event in a given time period	Probability of a given number of event occurrences using the Poisson distribution	The Poisson occurrence rate in the Poisson distribution	Uncertainty distribution in the Poisson occurrence rate
Length of crack growth after a given number of operational cycles	Probability of crack growth length using the Paris growth model	The material parameters in the Paris crack growth model	Uncertainty distributions in the material parameters in the Paris crack growth model
No failure occurring in a spacecraft in a mission	Probability of no failure occurring using an appropriate failure model including internal causes and MMOD impacts	Component failure rates, MMOD occurrence rates, material strength parameters	Uncertainty distributions in the failure rates, MMOD rates, and material strength parameters
Specific assurance checks and tests conducted successfully before a new launch of a spacecraft	Probability of a successful launch given successful assurance checks and tests using a Bayesian net model	Conditional probabilities in the Bayesian net relating the success probability to the checks and tests conducted	Uncertainty distributions in the conditional probabilities in the Bayesian net model
Measurements of deviations in function outputs of an operating system to check on its future life length	Probability of the time of system failure using the Weibull random process degradation model	Shape and scale parameters in the Weibull random process degradation model	Uncertainty distributions in the Weibull process shape and scale parameters

Revising a Risk Estimate with Zero Observed Failures Using a Poisson Model

Since it often occurs in risk estimation, it is informative to present the application of Bayes rule in revising a risk estimate with zero observed failures. As subsequent sections will illustrate, significant information can be obtained in revising the risk estimate by applying Bayes rule to revise an MMOD risk estimate with no observed failures as well as with failure and with observed anomalies. The information that is obtained is contrasted with the minimal information that is obtained in using classical, empirical statistics to analyze the observed data in which the initial (prior) risk estimate and its uncertainty are not used to obtain the revised risk estimate. The application of Bayes rule is presented when zero failures are observed and when a Poisson model is used for the occurrence of no failure. The Poisson model is a common model used in risk assessment in general and in MMOD risk assessment in particular.

For zero observed failures, using a Poisson model, Bayes rule (Equation (67)) becomes

$$f(\Lambda/0, T) = \frac{\exp(-\Lambda T) f(\Lambda)}{\int_{\Lambda'} \exp(-\Lambda' T) f(\Lambda') d\Lambda'} \quad (71)$$

where

$$\Lambda = \text{the Poisson occurrence rate for failure} \quad (72)$$

$$T = \text{the cumulative observed time period in which no failure occurred} \quad (73)$$

$$f(\Lambda) = \text{the uncertainty distribution (pdf) for } \Lambda \text{ before revision} \quad (74)$$

with observed data

$$f(\Lambda/0, T) = \text{the revised uncertainty distribution (pdf) for } \Lambda \text{ after observing} \quad (75)$$

0 failures in a cumulative time period T .

The term “cumulative observed time period” in the above refers to the total observed time over the sample set, e.g. set of spacecraft modeled as having the same failure occurrence rate. For a more general Poisson model where the Poisson occurrence rate is a function of time and location such as in a spacecraft mission trajectory then the term ΛT is replaced by the integral of the occurrence rate over the time and location of the trajectory, as previously indicated. For a spacecraft, the Poisson occurrence rate Λ can also be expressed as the occurrence rate per unit area integrated over the exposed area. These more detailed considerations are further treated in subsequent sections describing evaluations of the revisions of the MMOD estimated risk with different types of observation data.

Revising a Risk Estimate with an Observed Number of Events Using a Poisson Model

The previous application of Bayes rule for zero observed failures can be simply generalized to an observation of n events in a cumulative time period T where the events can be failures or anomalies over the sample set. For n events observed in a cumulative time period T

$$f(\Lambda/n, T) = \frac{\Lambda^n \exp(-\Lambda T) f(\Lambda)}{\int_{\Lambda'} (\Lambda')^n \exp(-\Lambda' T) f(\Lambda') d\Lambda'} \quad (76)$$

where

$$f(\Lambda/n, T) = \text{the revised uncertainty distribution (pdf) for } \Lambda \text{ after observing } n \text{ events in a cumulative time period } T. \quad (80)$$

(Note that the factorial $n!$ in the denominator of the Poisson probability cancels out since it appears in both the numerator and denominator of the right hand side of Equation (76).). Appropriate relationships need to be further incorporated if anomalies are only observed and a revised failure rate is to be estimated or if particular failure rate contributions are to be estimated. These considerations are further covered in the subsequent sections describing applications to the revision of MMOD risk estimates.

Using Fuzzy Observed Data

Fuzzy data occurs when the specific type, size or severity of the event is not accurately known. Examples include uncertainty regarding whether an event is a failure or not, regarding the precise number of events occurring, or regarding the cause of an event such as whether an MMOD impact is a micrometeoroid impact or an orbital debris impact. Fuzziness and uncertainty in the observed data can be viewed as simply another type of uncertainty and can be handled by appropriate probabilistic techniques. For fuzzy and uncertain data, the likelihood function $L(E; \nu)$ in Equation (67) is modified to represent the fuzzy and uncertain data. For uncertainty in the number of events observed the likelihood is modified to be a sum of probabilities over the possible events. For an uncertainty in the severity of the event the likelihood is modified to cover the possibilities of the different severities. For example, instead of an uncertain event being counted as a failure a mark is assigned to the event which represents the probability, or plausibility, of the event as being failure with 50% representing a non-informative representation. The count of failures then becomes a sum of marks. References 20 and 21 give further details on specific techniques for representing fuzzy and uncertain data in the likelihood function. Even though it is fuzzy and uncertain, such data can provide significant information in revising and reducing the uncertainty associated with an estimate. The applications given in subsequent sections for revising MMOD risk estimates cover the handling of fuzzy and uncertain data involving MMOD impacts. Subsequent sections will describe specific instances of handling of fuzzy observed data in the of updating and revision MMOD risk estimates

Comparing the Consistency of a Model Prediction with Observed Data

Before revising an estimate with observed data it is useful in many cases to initially compare the consistency of the estimate with its associated uncertainty with the observed data. The consistency is

evaluated by constructing the predicted probability of the observed data using the estimate and accounting for the uncertainty in the estimate. The predicted probability measures the consistency of the prediction with the observation. If the predicted probability of the observed data is small, e.g. <1%, then this indicates that the observed data is not consistent with the model predication. The model or the observed data then needs to be further evaluated for its applicability or possible modification. Depending on the finding, the estimate can then be revised with the observed data with the revised estimate and associated uncertainty incorporating the observed data.

Using probability and statistical relationships, the predicted probability of the observed data is formally evaluated using the likelihood $L(E; \nu)$ for the observed data E . The likelihood is given by Equation (68) where ν is the associated parameter value (or values) in the likelihood. Accounting for the uncertainty in the parameter value ν , the predicted probability of the observed data occurring $P(E)$ is

$$P(E) = \int_{\nu} L(E; \nu) f(\nu) d\nu \quad (81)$$

where $f(\nu)$ is the parameter uncertainty distribution (pdf). The multiplication by the uncertainty distribution $f(\nu)$ and the integration over all values ν is important since it accounts for the uncertainty in the parameter in determining the probability of the observed event.

For example, using the Poisson model the predicted probability $P(0)$ of no event occurring in given time period T is

$$P(0) = \int_{\Lambda} \exp(-\Lambda T) f(\Lambda) d\Lambda \quad (82)$$

where $f(\Lambda)$ is the uncertainty distribution (pdf) in the Poisson occurrence rate Λ . For example if the uncertainty distribution is represented by the lognormal then Equation (82) becomes

$$P(0) = \int_{\Lambda} \exp(-\Lambda T) \frac{1}{\Lambda \sigma \sqrt{2\pi}} \exp\left(-\frac{(\ln \Lambda - \mu)^2}{2\sigma^2}\right) d\Lambda \quad (83)$$

where μ and σ are the mean and standard deviation of $\ln \Lambda$. The integration can be carried out in a straightforward manner using various software.

In some applications, it is more relevant to compare the predicted probability of the observation being as large or larger than the observed value or as small or smaller than the observed value, whichever is relevant. This is the case, for example, where the observed data is the length of a crack or the time of failure of a piece of equipment. The relevant probability for the crack length observation is then the probability that the crack length is as large or larger than that observed for the given number of cycles. The relevant probability for the time of failure is accordingly the probability that the time of failure is less than or equal to that observed.

For the applications where the probability of the observation being as extreme as that observed, the likelihood for the observed data is replaced by the likelihood integrated or summed over the appropriate interval. For an observation being as small as that observed the relevant likelihood for an observed value which can be continuous such as time of occurrence is

$$L(E^-; v) = \int_{E' \leq E} L(E'; v) dE' . \quad (84)$$

For an observation which is discrete, such as the number of occurring events, the relevant likelihood is

$$L(E^-; v) = \sum_{E' \leq E} L(E'; v) . \quad (85)$$

Similar relationships apply for the relevant likelihood for an observation being greater than the value observed. For a continuous observed value

$$L(E^+; v) = \int_{E' \geq E} L(E'; v) dE' . \quad (86)$$

For a discrete observation

$$L(E^+; v) = \sum_{E' \geq E} L(E'; v) . \quad (87)$$

As before, the relevant predicated probability is then obtained by integrating over the parameter uncertainty distribution to account for the parameter uncertainty. The predicted probability $P(E^-)$ for an observation to be as small or smaller than that actually observed is

$$P(E^-) = \int_v L(E^-; v) f(v) dv \quad (88)$$

The predicted probability $P(E^+)$ for an observation to be as large or larger than that actually observed is

$$P(E^+) = \int_v L(E^+; v) f(v) dv . \quad (89)$$

For the relevant case, the predicted probability is determined and measures the consistency of the prediction with the observation. The use of these relationships in evaluating the consistency of predicted values with observed data for MMOD risk predictions will be given in subsequent sections.

Using Observed Data to Compare Different Models

When there are alternative, candidate models then the relationships in the previous section can be used to compare the consistencies of the alternative models with the observed data. And more importantly, the observed data can be used to revise the relative supports for the alternative models to obtain a revised consensus model and revised consensus estimate. The revised consensus model will appropriately account for the different consistencies of the model along with their uncertainties. As before, two alternative models and their individual predications or individual estimates are considered. Generalization to more than two alternative models is again straightforward.

As previously, before including the observed data for revision, let

$$P(M_1) = \text{the relative support for Model 1} \quad (90)$$

$$P(M_2) = \text{the relative support for Model 2} \quad (91)$$

where

$$P(M_1) + P(M_2) = 1. \quad (92)$$

Also as before let

$$f_1(y) = \text{the probability that the estimated quantity has value } y \text{ according to Model 1} \quad (93)$$

$$f_2(y) = \text{the probability that the estimated quantity has value } y \text{ according to Model 2.} \quad (94)$$

and

$$f(y) = \text{the consensus probability that the estimated quantity has value } y \text{ by combining the evaluations from both Models 1 and 2.} \quad (95)$$

where

$$f(y) = P(M_1)f_1(y) + P(M_2)f_2(y). \quad (96)$$

The probability distribution $f(y)$ is the consensus distribution before revising with the observed data.

To determine the revised supports for the models based on the observed data E let

$$P(M_1 / E) = \text{revised support for Model 1 incorporating the observed data } E \quad (97)$$

$$P(M_2 / E) = \text{revised support for Model 2 incorporating the observed data } E \quad (98)$$

Then using Bayes rule, i.e. using the conditional probability relationship,

$$P(M_1 / E) = \frac{P(E / M_1)P(M_1)}{P(E)} \quad (99)$$

$$P(M_2 / E) = \frac{P(E / M_2)P(M_2)}{P(E)} \quad (100)$$

where

$$P(E / M_1) = \text{probability of the observed data using (given) Model 1} \quad (101)$$

$$P(M_1) = \text{support for Model 1 before the observed data} \quad (102)$$

$$P(E / M_2) = \text{probability of the observed data using (given) Model 2} \quad (103)$$

$$P(M_2) = \text{support for Model 2 before the observed data} \quad (104)$$

Finally using the total probability relationship

$$P(E) = P(E / M_1)P(M_1) + P(E / M_2)P(M_2) . \quad (105)$$

The revised supports for the individual models then become

$$P(M_1 / E) = \frac{P(E / M_1)P(M_1)}{P(E / M_1)P(M_1) + P(E / M_2)P(M_2)} \quad (106)$$

$$P(M_2 / E) = \frac{P(E / M_2)P(M_2)}{P(E / M_1)P(M_1) + P(E / M_2)P(M_2)} \quad (107)$$

Focusing on the numerator of Equation (106), the revised support for Model 1 is thus proportional to the consistency $P(E / M_1)$ of the observed data with the prediction of Model 1 times the initial support

$P(M_1)$ of Model 1. Similarly for Model 2, the revised support is proportional to the consistency of the observed data with the prediction of Model 2 times the initial support for Model 2. The denominator in the equations is the normalizing factor to have the revised supports add to 1. The consistency of a model prediction with the observed data is that determined in the last section. For example, for Model 1

$$P(E/M_1) = \int_v L_1(E;v) f_1(v) dv \quad (108)$$

where

$$L_1(E;v) = \text{likelihood of the event according to Model 1} \quad (109)$$

$$f_1(v) = \text{Model 1's uncertainty distribution (pdf) for the parameter } v. \quad (110)$$

Similar relationships hold for the consistency of Model 2. The subscripts in the relationships refer to the particular model involved. Importantly, the revised supports account for the uncertainties in the model predictions which is essential in assessing the model predictions and supports.

With regard to applications, for most cases, the likelihood function is the same for both Model 1 and Model 2, for example both Model 1 and 2 using the Poisson model to describe the probability of occurrence of the observed event. The differences between the models are then the different estimated values and uncertainty distributions in the Poisson occurrence rates. As a special case, which often occurs, It is important to note that for equal initial model supports, $P(M_1) = P(M_2)$ Equations (106) and (107) simplify to

$$P(M_1/E) = \frac{P(E/M_1)}{P(E/M_1) + P(E/M_2)} \quad (111)$$

$$P(M_2/E) = \frac{P(E/M_2)}{P(E/M_1) + P(E/M_2)}. \quad (112)$$

The revised model supports are thus simply the relative consistencies of the model predictions with the observed data. The more general relationships for the revised model supports given by Equations (106) and (107) account for the different initial supports for the models $P(M_1)$ and $P(M_2)$ according to their different technical bases and justifications.

Accounting for the revised model supports the revised probability for the estimated value y of the quantity of interest is then

$$f(y/E) = P(M_1/E)f_1(y/E) + P(M_2/E)f_2(y/E) \quad (113)$$

where

$$f(y/E) = \text{the revised consensus probability distribution accounting for the observed data} \quad (114)$$

$$f_1(y/E) = \text{the revised probability distribution for Model 1 accounting for the observed data} \quad (115)$$

$$f_2(y/E) = \text{the revised probability distribution for Model 2 accounting for the observed data.} \quad (116)$$

The revised probability distributions use Bayes rule with the likelihood for the observed data and the prior distribution as discussed in the section on revising a model estimate with observed data. For example, as applied to revising an MMOD risk estimate the observed data can be the observed number of MMOD impacts, or the observation of no impacts, in a given time period over a given collection area. The parameter ν will then be the appropriate occurrence rate for the impacts in the Poisson model. The two uncertainty distributions for ν will be those in the individual models. The quantity of interest with possible value y will be the failure occurrence rate, which may be different from the occurrence rate for the observed impacts which may not involve failure. The relationship between the occurrence rates will be constructed based on the information related to the observed impacts. The sections describing MMOD applications cover these considerations in more detail.

Uses of Quantified Uncertainties in Decision-Making

The previous sections described the uses of quantified uncertainties in determining the precision and accuracy of an estimate, in comparing and combining different model estimates, in determining the uncertainty contributors to a result uncertainty, and in revising an estimate and its uncertainty with observed data. Quantified uncertainties also can be used in the following ways in decision-making:

1. Assessing whether the uncertainty in an estimated value is so large or has such a lack of characterization as to require re-assessment and error reduction before the estimate can be meaningfully used in decision-making.
2. Using an appropriate estimated value from the uncertainty distribution to provide a given confidence in a numerical requirement being satisfied.
3. Using the uncertainty distribution in a statistical-decision framework to formally determine an estimate or action that results in minimum loss and maximum benefit.

The consideration involved in each of these uses will be discussed in the following.

Assessing Whether the Uncertainty Allows an Estimate to be Meaningfully Used

The type of assessment can be termed a sanity check on the uncertainty. When the uncertainty bounds are assessed, the conclusion may be that the bounds are so large or are so ill-defined that the estimate cannot meaningfully be used in the intended application. A reassessment and reconstruction of the bases for the estimate and associated uncertainty and uncertainty contributors may then be needed.

The situation can occur when the estimate needs to have a given tolerance or given specified bounds for the application. As discussed for the next type of assessment in determining the confidence, the assessed uncertainty bounds may be so large that a specified requirement or goal is not satisfied with adequate confidence. The use of the estimate may then be held in abeyance until the uncertainty is reduced. A value of an uncertainty quantification is that the size of the uncertainty becomes evident which would not be recognized otherwise thereby preventing an erroneous use of the estimate for decision-making.

Determining the Confidence in the Satisfaction of a Requirement

When an estimate is compared with a numerical requirement then the confidence in satisfying the requirement is an important piece of information. The confidence is the probability that the true value actually satisfies the requirement. Even though the estimated value satisfies the requirement the probability that the true value satisfies the requirement may be unacceptably small because of the large uncertainties associated with the estimated value. Having adequate confidence in a requirement being satisfied is especially important if not satisfying the requirement can have significant consequences. It is also important if the actual value can be significantly higher or lower than the estimated value resulting in significant, different consequences when the actual value is lower or higher than the estimate. This can occur if an estimated risk value is compared with an acceptability value. If the estimated risk value has high uncertainty and is higher than the acceptability value then the mission or activity may be cancelled even though the actual risk value is much lower than the estimated value. On the other hand, if the estimated risk value has high uncertainty and is lower than the acceptability value then the actual risk value can be significantly higher than the acceptability value resulting in high risk being incurred. By quantifying the uncertainty associated with the estimate the probability that the true value satisfies the requirement is determined given the estimated value does.

One approach that is used in addressing the uncertainty with an estimated value is to define a separate threshold requirement and a separate goal. As applied to a spacecraft, for example, the threshold requirement defines the maximum risk that is acceptable and the goal defines a lower risk value to be the target in the longer term. In this case, the confidence in satisfying the threshold is an equally important piece of information if the consequences of not satisfying the threshold requirement can be significant. The importance of the confidence also applies to satisfying the goal where the confidence may be lower in value. Without information on the confidence one is basically going forward with no assurance that the true value satisfies the goal. Furthermore, without assessing the uncertainty it is not clear whether the estimated value is conservative or nonconservative or something in between.

An acceptability requirement can be two-sided or one-sided. For a two-sided requirement both a lower bound and upper bound are specified and the estimated value needs to lie between the bounds. For a one-sided bound either a lower bound or upper bound is specified. For a lower bound requirement the estimated value needs to be above the lower bound and the confidence is the probability that the true value is actually above the requirement value. For an upper bound requirement the estimated value needs to be below the upper bound and the confidence is the probability that the true value is actually below the upper bound requirement. For two-sided requirement bounds the confidence is the probability that the true value actually lies between the two bounds. The following considerations apply

to the confidence associated with the true value satisfying an upper bound. The considerations also apply to satisfying a lower bound or a two-sided bound with appropriate confidences used.

The confidence in satisfying a requirement can be specified or can be determined by determining the probability associated with the maximum bound of the uncertainty distribution (the cdf) that satisfies the upper bound requirement. By comparing the median value to the requirement there is a 50% confidence that the true value satisfies the requirement and a 50% confidence that it does not. The problem is that the associated uncertainty taken into account. If there can be significant consequences that the true value is significantly higher than the upper bound value then the use of the median value can be problematic. The median value can be used if the uncertainty associated with the estimate is determined also to be acceptable, e.g. if the spread of the 5% bound to 95% bound is determined also to be acceptable. However, this is equivalent to using an upper bound value from the uncertainty distribution to compare to the upper bound requirement.

The mean value of the uncertainty distribution can be used to compare with an upper bound requirement. An argument for the use of the mean value is that the mean value minimizes the loss from the true value deviating from the estimated value for an assumed quadratic loss function. The use of loss functions in a statistical-decision framework is discussed in the next section. A problem with this argument is the assumption that there are equal consequences of true value being higher than the estimated value or being lower than the estimated value. Use of the mean value in turn implies that there are equal consequences of the true value being higher than the upper bound requirement or being lower than the upper bound requirement. This is generally not necessarily true since the upper bound requirement only focuses and controls the risk of the true value being higher than the upper bound requirement. The use of the mean value is thus more suited in using an estimated value to compare with a two-sided requirement. And importantly, the use of the mean value furthermore does not provide any specific confidence of the true value satisfying the upper bound requirement. The uncertainty distribution may be so negatively skewed the confidence can be significantly below 50%. This does not generally occur in risk estimates but it depends on the risk model and input parameter uncertainty distributions.

Particularly in risk assessments, it is sometimes argued that the mean value corresponds to approximately the upper 75% bound value of the uncertainty distribution. Thus if the mean value satisfies the upper bound requirement then this is associated with a confidence of approximately 75% of the true value satisfying the requirement. If this correspondence is checked with the uncertainty distribution of the estimate and if an approximate 75% confidence is acceptable then this can be an acceptable approach. However, in many cases it is more direct to use an appropriate upper bound value from the uncertainty distribution to compare with the requirement.

Several considerations enter in selecting a confidence level to be associated with satisfying a requirement. These involve the strictness and control desired in satisfying the requirement, the consequences of any violation of the requirement, and the robustness of the uncertainty distribution used to determine the confidence level. The stricter the requirement is to be satisfied the greater the confidence level to be used in satisfying the requirement. A 90% or 95% confidence level is generally used for such a strict requirement. This strictness also applies when there are significant consequences

of any violation of the requirement. In deciding upon whether to use a high confidence level, consideration needs to be given to the robustness of the uncertainty distribution used for the confidence level determination. The tails of the uncertainty distribution, e.g. the highest and lowest 10% tails, can be sensitive to the detailed shapes and tails of the distributions used to determine the result distribution. If a high confidence level is to be used then associated requirements or guidelines need to be placed on the uncertainty distributions and analyses used to determine the resulting uncertainty distribution used for the confidence level determination.

For lesser strictness and control, but still providing adequate confidence, a confidence level of 75% can be used. This confidence level corresponds to the approximate confidence level associated with the mean value of positively skewed uncertainties distributions, such as the lognormal, with larger uncertainty, e.g. with an error factor of 10. As previously indicated this correspondence occurs for various probabilistic risk assessments performed. Using a 75% confidence level also does not involve the sensitivities associated with the tails of the uncertainty distribution. Thus, a 75% confidence level is a practical choice when there can be variations in the uncertainty distributions used and the uncertainty analyses applied.

The above considerations are intended to be guidelines and the specific confidence level selected depends on the particular application. A tolerance level can also be associated with a confidence level value such as 5%. When a confidence level is not attained as part of the requirement then the requirement is not satisfied due to the uncertainty associated with the estimate. Using the techniques described in the document, the uncertainty contributors could then be identified and further information gathered and analyses performed to reduce the dominant uncertainty contributors.

Accounting for the Uncertainty in a Statistical-Decision Framework

In making a decision, uncertainties can be formally taken into account using a statistical-decision framework. In this framework, a loss-function is defined which gives the loss for a given action for a possible value of the uncertainty quantity. Benefits are also included as a negative loss. The action can also be the selection of an estimated value for the quantity which accounts for the uncertainty in the quantity. The loss function for a given action is then integrated over the uncertainty distribution for the quantity to give the total loss for the action. The action which minimizes the total loss is then selected as the optimal action. The use of a loss function and integration over the uncertainty distribution is theoretically the optimal way of accounting for uncertainties to identify the action which minimizes loss. The practical problems involve defining the form of the loss function and transforming losses and benefits to a same scale. However, useful insights can be gained from the optimal, minimum loss actions that are determined for various types of loss functions.

Formally, in a statistical-decision framework, the loss for a given action is given by (22,23):

$$L(a) = \int_y l(a, y) f(y) dy \quad (117)$$

where

$$L(a) = \text{the total loss for the action } a \text{ accounting for the uncertainty in the quantity} \quad (118)$$

$$l(a, y) = \text{the loss for the action } a \text{ for a particular value of the } y \text{ of the quantity} \quad (119)$$

$$f(y) = \text{the uncertainty distribution (pdf) for the quantity.} \quad (120)$$

The integral in Equation (117) is over all possible values of y . Also, let

\tilde{a} = the optimal action which minimizes $L(a)$, i.e.,

$$L(\tilde{a}) = \min_a \int_y l(a, y) f(y) dy \quad (121)$$

As a particular case let

$$a = \text{an estimated value from the uncertainty distribution} \quad (122)$$

$$l(a, y) = K(y - a)^2 \quad (123)$$

where

$$K = \text{some scaling constant.} \quad (124)$$

Then the mean value of the uncertainty distribution is the optimal estimated value \tilde{a} which minimizes the loss due to the true value of the quantity deviating from the estimated value. This applies to any uncertainty distribution. As previously indicated this is useful information when the losses of being above or below the estimated value to use are approximately equal and when they significantly increase with greater deviations of the true value from the estimated value to use. As was also indicated for risk assessments the mean value of the uncertainty distribution often corresponds to approximately the 75% upper bound.

The median value is the optimal estimate when the loss function is proportional to the absolute distance the true value is from the estimated value, i.e.,

$$l(a, y) = K|a - y| \quad (125)$$

where K is a proportionality constant and the vertical bars denote absolute value. Thus the median value is an optimal estimate for a two-side loss which is linear in the size of deviation of the true value from the estimated value. As was previously indicated, the median value is also a more robust estimate and is not sensitive to the tails of the uncertainty distribution

As opposed to two-sided requirements, one- sided upper bound requirements correspond to loss functions that have greater loss when the true value is greater than the estimated value as opposed to being less than the estimated value. For a loss function defined as

$$l(a, y) = K_0(y - a) \quad y \geq a \quad (126)$$

$$l(a, y) = K_1(a - y) \quad y < a \quad (126)$$

then the optimal estimate \tilde{a} is

$$\tilde{a} = \frac{K_0}{K_0 + K_1} \quad (127)$$

Thus when $K_0 > K_1$ the optimal estimate is an upper bound of the uncertainty distribution.

The above results are interesting in that they show that optimal estimates from a statistical-decision standpoint correspond to the estimates that were obtained in the previous sections which were application and operational oriented. One additional, basic result is of interest which is obtained from a utility theory and decision-theoretic standpoint and which involves the decision as to whether to launch a spacecraft or not. Let

$$L = \text{the loss from failure of the launch} \quad (128)$$

$$B = \text{the benefits from success of the launch} \quad (129)$$

For more involved models there could be different losses from different types of failures or partial-failures and different benefits from different partial successes. However, the above, simpler model demonstrates the information obtained from utility theory and the difficulty involved. For the simpler model let

$$F = \text{the probability of a launch failure.} \quad (130)$$

To account for uncertainties in the probability of failure is the average value of the associated uncertainty distribution. Assuming that losses and benefits can expressed on a common scale, which is the differently, the net benefit B_{net} of the launch is

$$B_{net} = (1 - F)B - FL \quad (131)$$

Thus, there will be positive net benefits when $B_{net} > 0$ or when

$$(1 - F)B - FL > 0 \quad (132)$$

or

$$F < \frac{B}{B + L} \quad (133)$$

which can also be expressed as

$$F < \frac{1}{1 + L/B} \quad (134)$$

This relationship can be used to relate a upper bound requirement on failure probability to an equivalent implied value for loss benefit ratio . For example, requiring

$$F < 0.01 \quad (135)$$

corresponds to

$$L/B = 99 \quad (136)$$

i.e. the losses are viewed as being 99 times the benefits . A failure probability requirement of 0.001 corresponds to the losses being viewed as being 999 times the benefits thereby requiring a more stringent failure probability requirement. It can be informative to compare different failure probability requirements that have been imposed to obtain the implied loss-benefit ratios. These could be compared for a relative scaling of the benefits and losses.

Examples of MMOD Risk Uncertainty Analysis and Data Analysis

The following sections give examples of MMOD risk uncertainty analysis and data analysis involving comparison of the lognormal versus gamma uncertainty distributions, comparing the predictions of two MMOD models with observed data of no MMOD occurrences in a given time period, updating the individual model MMOD predictions with the data of zero observed occurrences, combining the individual model predictions to obtain a consensus model prediction, and dynamically updating MMOD predictions with yearly observed periods of no MMOD occurrences or of a given number of occurrences. Example flow Charts and evidence networks are given after the examples which illustrate the topics and issues that need to be addressed in designing MMOD data collection and analysis efforts. Finally, example coding is given for the WINBUGS software package that was used for the simulations. Even though the focus is on MMOD applications, the approaches have general application.

Comparison of Lognormal Versus Gamma Uncertainty Distributions

Figures 5 and 6 compare the distributions for lognormal and gamma uncertainty distributions for the same specified median and 95% error factor where the 95% error factor is defined as the 95% upper bound divided by the median. A median of 0.05 and a 95% error factor of 10 are specifically shown which for example represents a predicted MMOD failure rate of 0.05 per year with an error factor of 10, but can apply to any case. The figures show that the uncertainty distributions (cdfs) are very close and practically indistinguishable for probabilities of 0.5 and greater due to the specification of the same median value and 95% error factor value. For probabilities less than 0.5 the probabilities diverge with the gamma distribution having a higher probability of the quantity being less than a given value (or alternatively for the same probability the gamma giving a lower quantity value). This divergence is due

to the gamma distribution giving higher probabilities to low values approaching zero. Figures 7 and 8 show the posterior uncertainty distributions for the lognormal and gamma distributions that have been updated and revised with observed data of no failures occurring in 5 and 10 years, respectively. The figures again show the distributions are practically indistinguishable for probabilities of 0.5 and greater due to the prior distributions having the same median and 95% error factor. This indistinguishable behavior for probabilities of 0.5 and greater generally occurs for other cases. The implications are that the gamma and lognormal are indistinguishable within a few percent for larger uncertainties if they have the same medians and 95% error factors and if the median and higher bounds are of principle interest. Since the parameters of the gamma distribution are simply updated with observed data then this can be a motive for using the gamma distribution to describe the uncertainty distribution. When there is a question then both the lognormal and gamma can be used and be averaged.

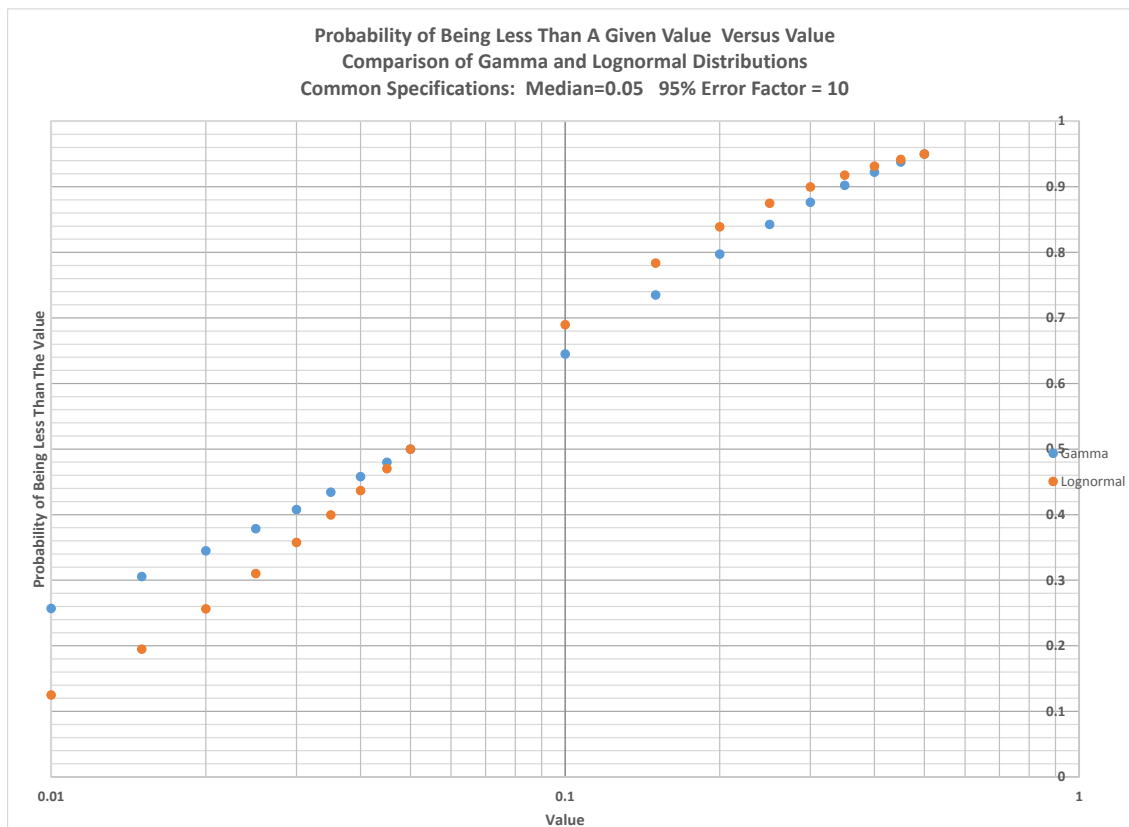


Figure 5. Lognormal Versus Gamma Uncertainty Distributions for the Same Median and Error Factor

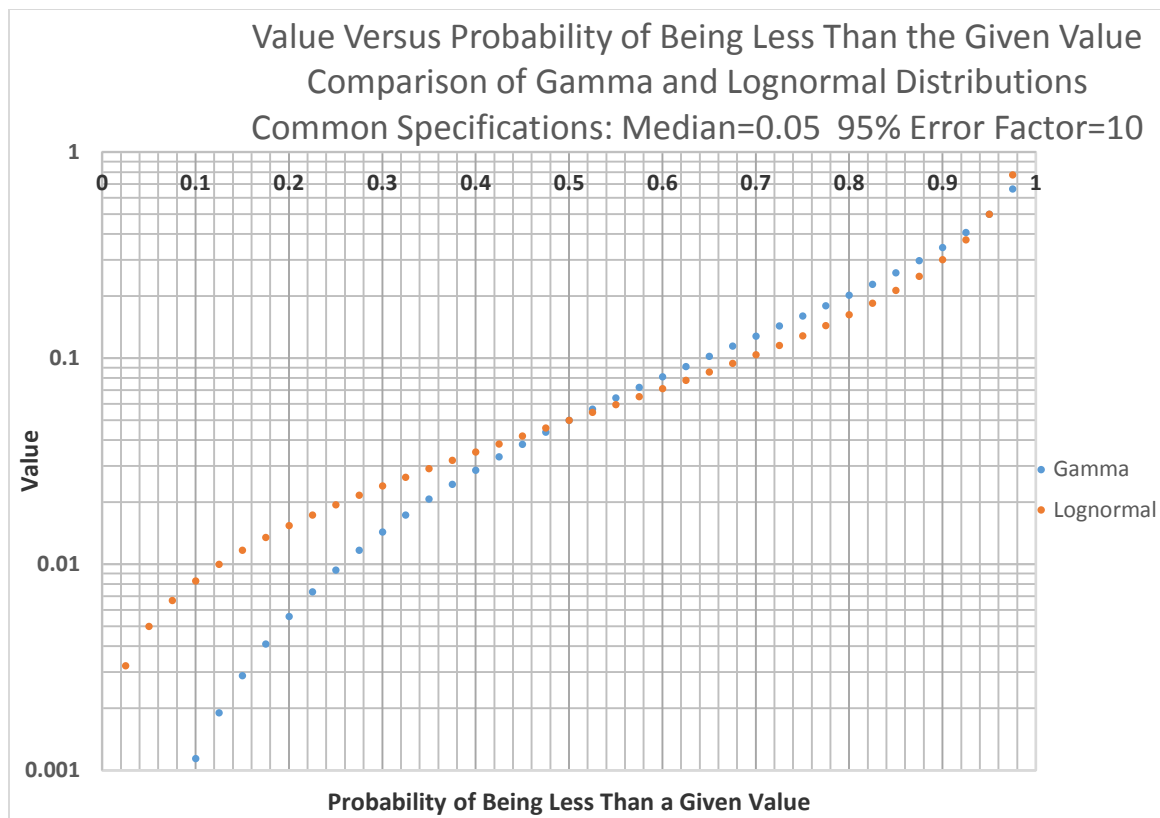


Figure 6. Alternative Presentation of Lognormal versus Gamma Uncertainty Distributions

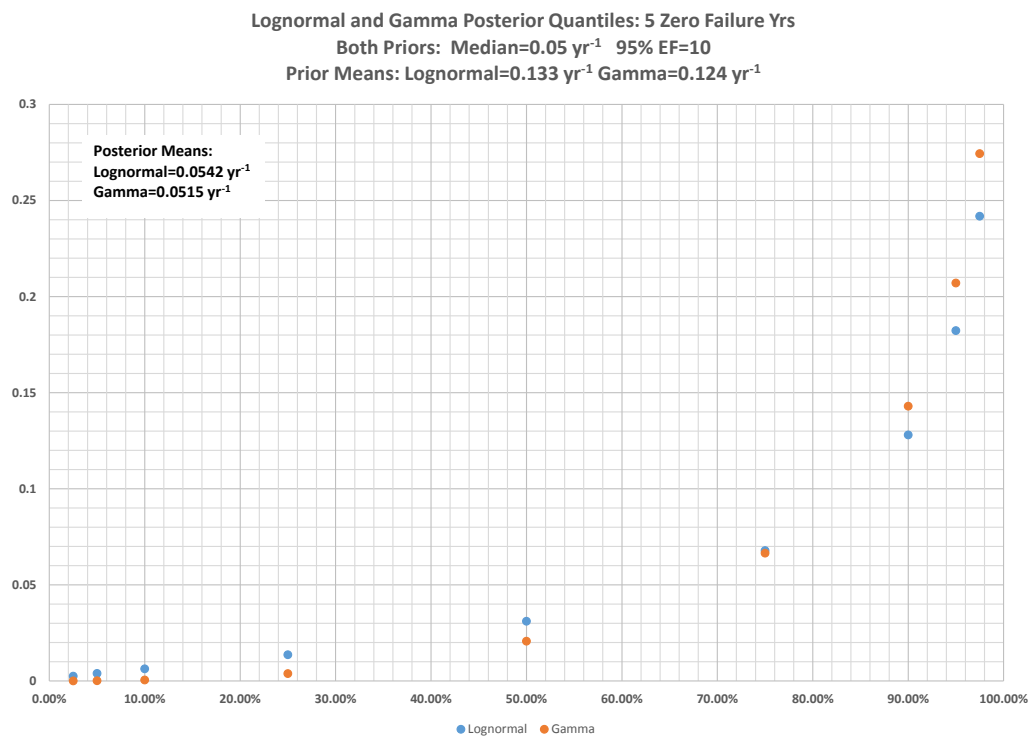


Figure 7. Lognormal versus Gamma Posterior Distributions for 5 Zero Failure Years

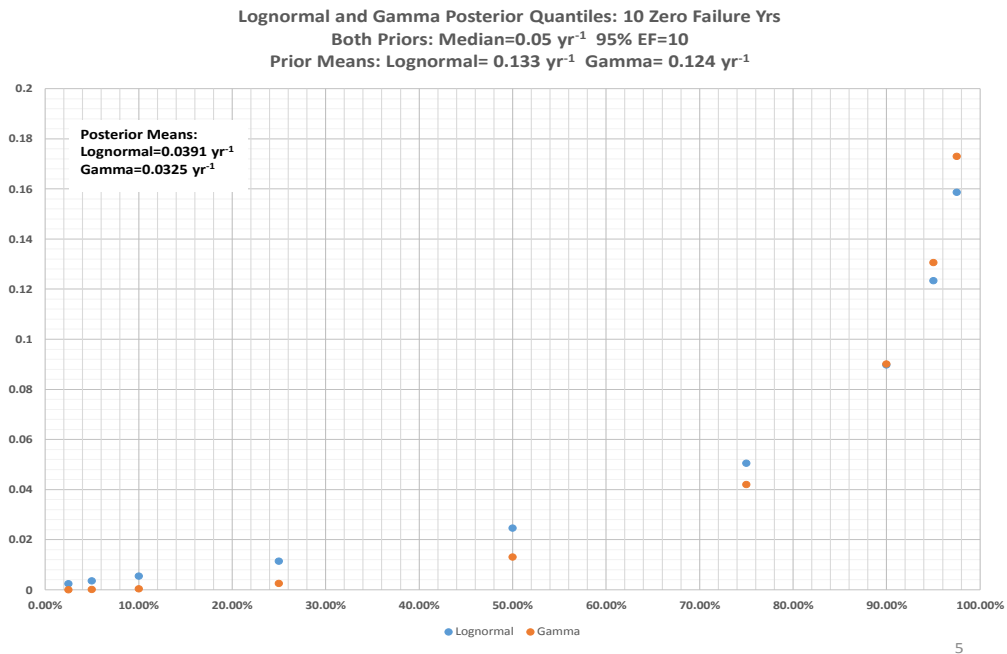


Figure 8. Lognormal versus Gamma Posterior Distributions for 10 Zero Failure Years

Comparing and Updating Model Predictions with Observed Data of No MMOD Occurrences

The following graphs illustrate the information that can be provided in comparing and updating MMOD predictions from different models using observed data. The figures are only illustrative but they show the type of information that is obtained from data. The first set of figures uses observed data with no MMOD occurrences which importantly illustrates the useful information that is obtained even when no MMOD occurrences are observed. The model predictions are the probabilities of no MMOD occurrences as observed. The MMOD occurrences are usually failure causing occurrences but can be of any type and severity. For accurate comparisons and updating, the model predictions should be for the MMOD type as measureable by the detection equipment. Differences can be modeled in the likelihood as previously discussed in handling fuzzy information. Differences can also be included in expanding the uncertainty range associated with a predictions. Order of magnitude uncertainty ranges (error factors) are associated with the predictions in the figures which can account for the uncertainty contribution arising from differences between the predictions and measurements.

As previously described, the model prediction of no MMOD occurrences uses the Poisson model with the model-predicted MMOD occurrence multiplied by the area of detection.(Equation (82)). Using Bayes rule, the Poisson likelihood and the associated uncertainty distribution are then combined to obtain the updated, revised MMOD occurrence rate and uncertainty distribution (Equation 71). A gamma distribution with a specified median and 95% error factor is used for the uncertainty distribution of the predicted MMOD occurrence rate. use of the lognormal distribution with the same median and 95% error factor gives results that closely follow the gamma results. The figures show the model comparisons and updates versus different observed time periods of no MMOD occurrence, where the time period can be for one observation or can be the cumulative time for a set of observations.

Figure 9 shows the updated supports, or confidences, for two individual model predictions of no MMOD occurrences with observed data of no occurrence in given time periods. An initial supports of 50% is used for each model. A high predicted mean occurrence rate of 0.05 per year is used for one model which is termed the model which includes the high density flux contribution. A predicted mean occurrence rate of 0.005, which is one order of magnitude lower, is used for the second model which is termed the model which excludes the high density flux contribution. Both predicted occurrence rates have a 95% error factor of 10. The figure shows that as the observation period of no MMOD occurrences increases then the support for the model initially predicting the higher occurrence rate of 0.05 per year continually decreases. At the same time the support for the model predicting the lower occurrence rate of 0.005 per year increases. These results are illustrative of the revision and updating of the confidence in an individual model according to the degree of consistency of the model prediction with observation. As a counter, later slides will show the support decreasing in the model with the lower occurrence rate with observation of MMOD occurrences.

Figure 10 shows the updated mean MMOD occurrence rate for two individual model predictions and the consensus predicted occurrence rate accounting for the updated support (confidence) for each model. The mean occurrence rates are taken from the uncertainty distribution of the occurrence rate. The initial predicted mean occurrence rates for the two models are the same as given in the previous figure. The names of the models reflect a reference to specific MMOD models and are not of significance to the results. What the graphs show is that the initially high predicted occurrence rate of 0.05 per year with a large error factor is significantly reduced and revised to a lower predicted occurrence rate as the period of no observed occurrences increases. The lower occurrence rate of 0.005 per year is not significantly revised since the observation is consistent with the prediction. The consensus model is a compromise between the two models and approaches the lower occurrence rate model with an increasing period of no observed occurrences.

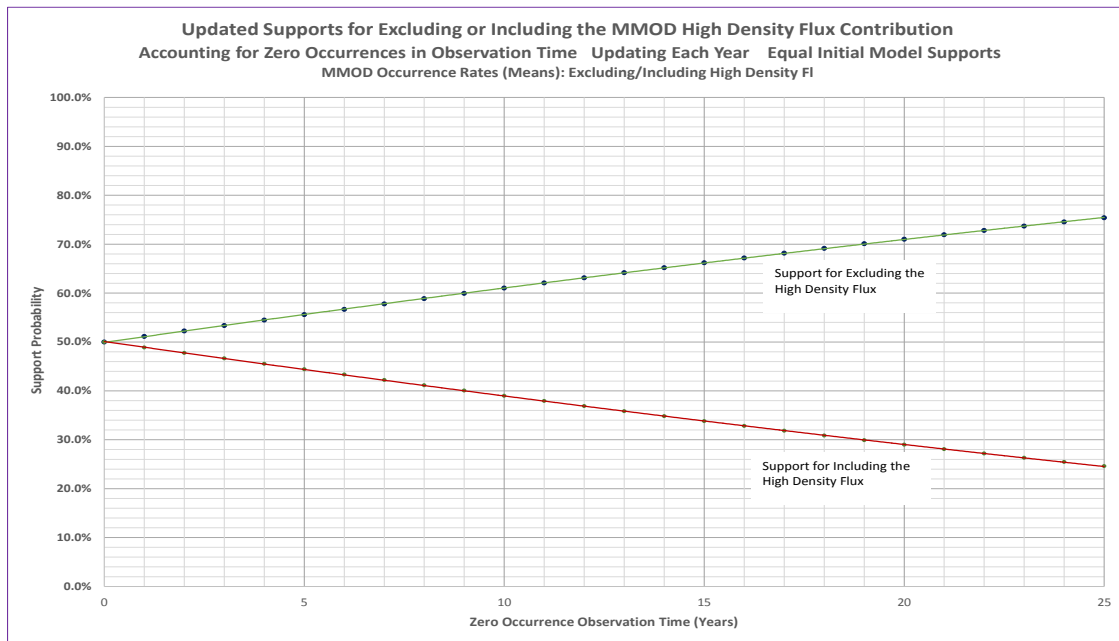


Figure 9. Updated Supports for Two Alternative MMOD Models versus Observed Time of Zero Occurrences

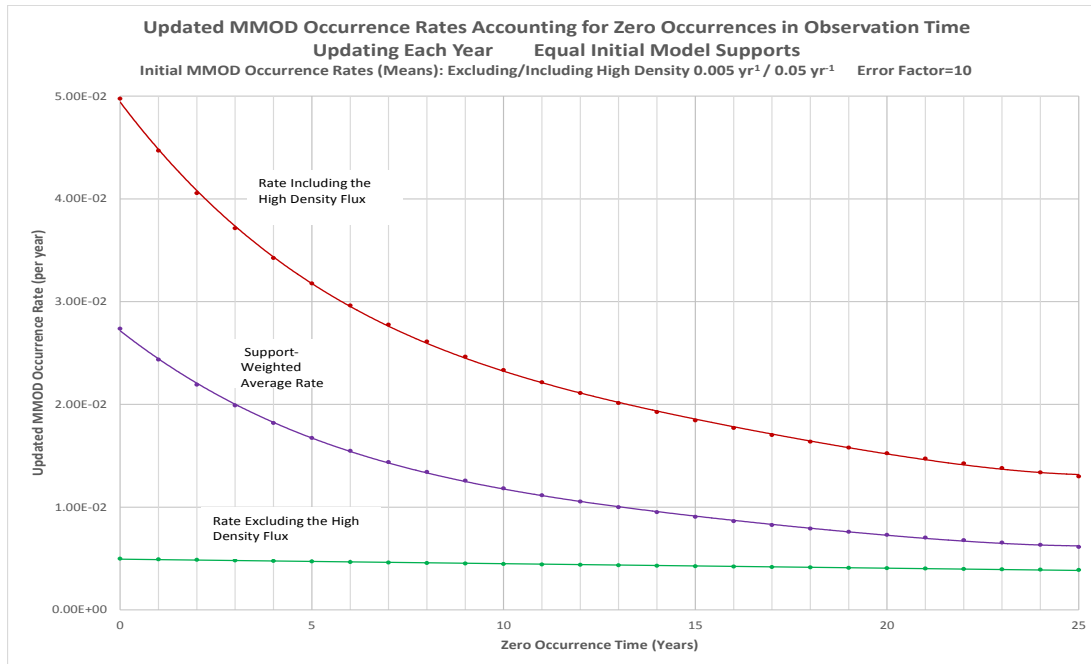


Figure 10. Updated MMOD Mean Occurrence Rates for Two Alternative Models versus Observed Time of Zero Occurrences

Figure 11 shows the updated mean MMOD occurrence rate versus the time of no observed MMOD occurrences comparing the interpretation of the initial central estimate of the occurrence rate as the mean versus the median.. The comparison of the mean versus the median is shown for the two models, including the high density flux contribution and excluding the high density flux contribution. A 95% error factor of 10 is used for each of the initial occurrence rate. A gamma distribution was used for the uncertainty distribution. The use of the lognormal distribution as the uncertainty distribution with the same initial occurrence rates and error factors gave relatively small differences. The figure shows that because of the positive skewness in the gamma or lognormal, using the central estimate as the median value causes the associated mean value to be higher by roughly a factor of 2.5. This difference is small compared to the error factor and the difference decreases as observed data is used to update the results, which all become more aligned with the observed data. The results of the model excluding the high density contribution shows less change with the data updating because of its greater consistency with the observed data.

Figure 12 shows the factor reduction in the updated mean occurrence rate with no observed MMOD occurrences for the high density model with an initial mean occurrence rate of 0.05 yr⁻¹ and for two 95% error factors. The factor reduction in the updated mean occurrence rate is greater for an error factor of 30 as compared to 10 and in general the factor reduction is greater the greater the error factor. This illustrates the capability of the data updating process to compensate for greater uncertainty in the estimated result to align with observed data. This capability will also be illustrated when non-zero occurrences are observed.

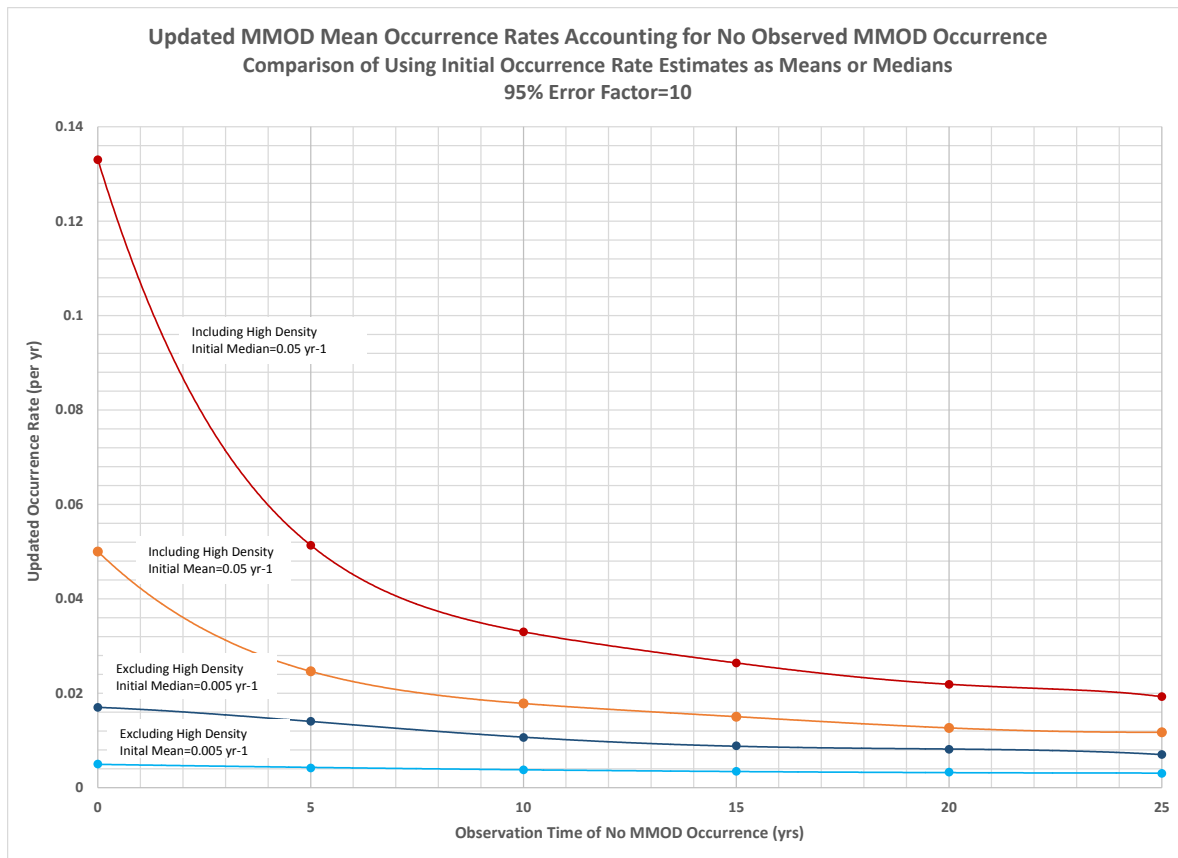


Figure 11. Updated Mean MMOD Occurrence Rates versus Observed Time of No Occurrence Using a Mean or Median as the Initial Occurrence Rate

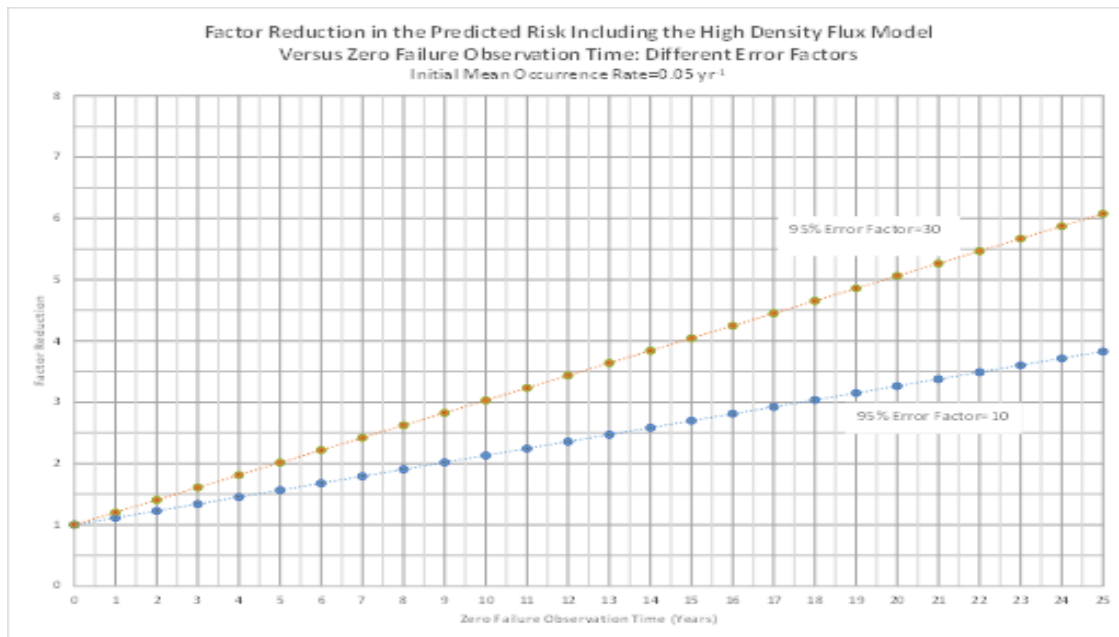


Figure 12. Factor Reduction in the Updated Mean MMOD Occurrence Rate for the High Density Model for Different Error Factors versus Observed Time of No Occurrence

Figures 13 and 14 are final illustrations of the evaluations of information that can be provided by observed data with no MMOD occurrences. Figure 13 shows updated MMOD occurrence rates for the model with no high density flux contribution versus time period of no MMOD occurrence. An initial median occurrence rate of 0.005 yr^{-1} is again assumed. The figure shows the updating for two 95% error factors used, a factor of 10 and a factor of 30. Both the updated mean occurrence rate and updated median occurrence rate are shown. As shown in the figure, the greater the revision in the median and mean occurrence rate to be consistent with the observed data. Since the median is significantly smaller than the mean and is in a different range, the change does not appear as significant as for the mean. Both the mean and median are converging with a greater time period of no occurrence. This illustrates the effect of the smoothing out effect and converging effect of observed data.

Figure 14 illustrates a more general design-of-experiments study that to determine the observation time period without an MMOD occurrence needed to decrease the support of an MMOD model versus its predicted MMOD occurrence rate. The decrease in the odds of support is defined as the probability of the model not being applicable divided by the probability of the model being applicable. The figure is applicable for any model with any given predicted MMOD occurrence rate. As seen, for a given odds of decrease, such as given by the upper red line for a decrease of a factor of 2, a longer time period is needed the smaller the predicted occurrence rate. Also from the figure, for a given predicted occurrence rate a longer time period is needed the greater the odds of decrease, such as going from a factor of 2 decrease to a factor of 100 decrease. The figure is somewhat deceptive in the many orders of magnitude shown on the graph which can give a large order of magnitude increase in the time for a small physical increment on the graph. Smaller regions of the graph can be expanded in the vicinity of the predicted occurrence rates of interest to focus on specific studies. The graph is also general in that, as indicated, the occurrence rate and time period can be of any pair of consistent unit, such as per year and years or per hour and hours. Similar graphs can be obtained for changing the odds of support of a given model, either increasing it or decreasing it, by given values versus the number of MMOD occurrences observed in a given time period.

Comparing and Updating Model Predictions with Observed Data of an MMOD Occurrence

The evaluations with observed MMOD occurrences in given time periods are the same as for no observed occurrences with the only difference is that the Poisson likelihood is used for the given number of observed occurrences in the given time period (Equation (76)). As simple illustrations of the dynamic updating when occurrences are observed, Figures 15 and 16 illustrate example updates with one observed MMOD occurrence in a given time period. The same two alternative models are again evaluated as in the previous evaluations, a model including a high density flux contribution with an initial median occurrence rate of 0.05 per year and a model excluding the high density flux contribution with an initial median occurrence rate of 0.005 per year. Both have 95% error factors of 10. Figure 15 shows the updating tracking when the updating is done every 5 years and an MMOD occurrence is observed in the first 5 years then no occurrences are observed for the remaining observation periods. Figure 16 shows the updating tracking when the updating is done every year and an MMOD occurrence is observed between the 8th and 9th year then no occurrences are further observed. The cumulative observed MMOD occurrences are shown in the bottom step function in Figure 16. Using the property of the Poisson if only a given time period is observed such as 15 years then the values of the curves at 15 years would be the result. The more detailed tracking gives the dynamic behavior of the updating and revised prediction. As observed all model results converge with increasing observation time. The detailed uncertainty distributions that are also obtained give all the other characteristics of the results.

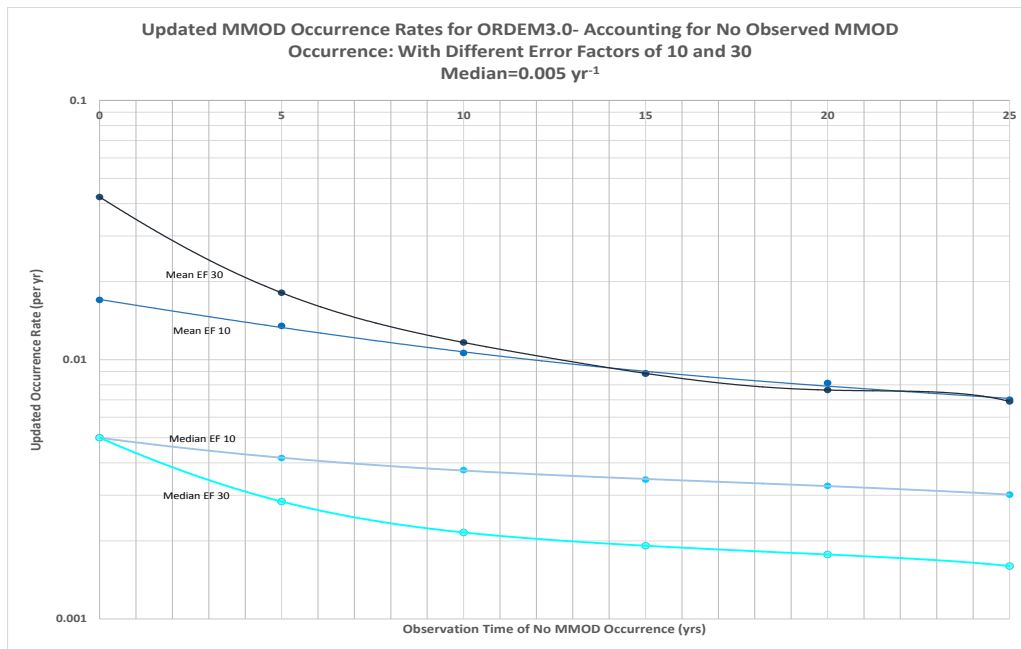


Figure 13. Updated Mean and Median MMOD Occurrence Rates for Different Error Factors versus Observed Time of No Occurrence

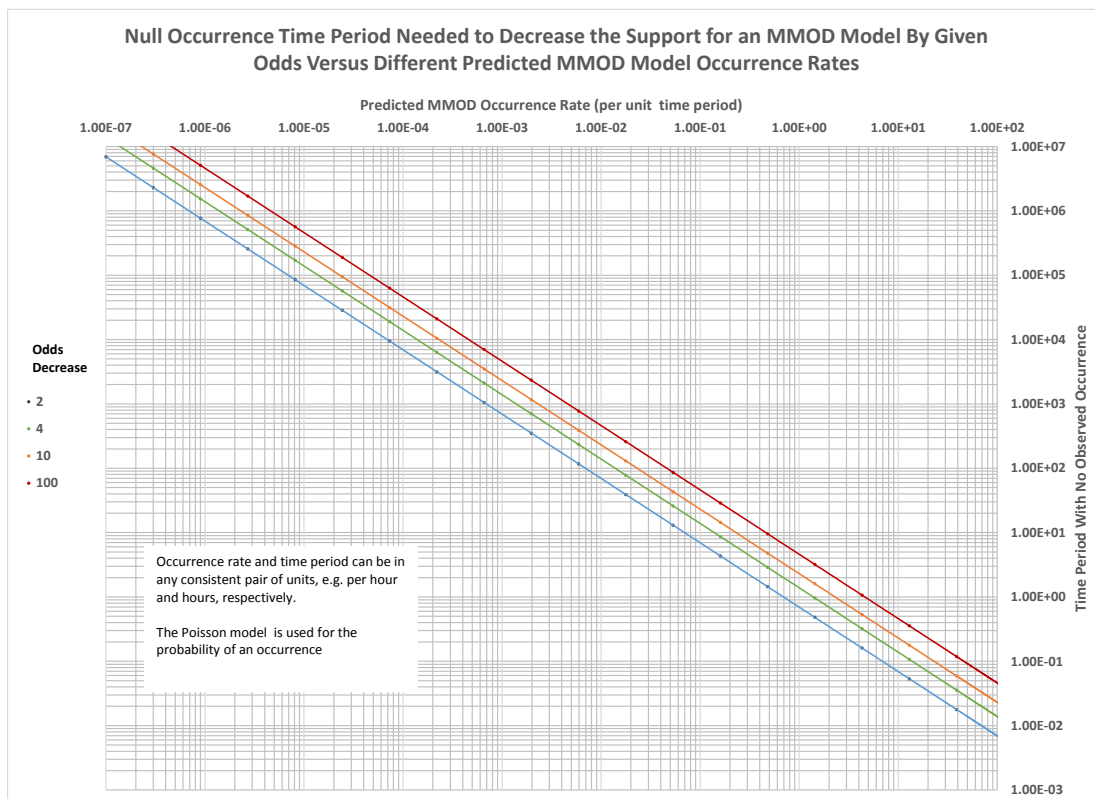


Figure 14. Observed Zero Occurrence Time Needed to Decrease the Support of an MMOD Model by a Given Odds

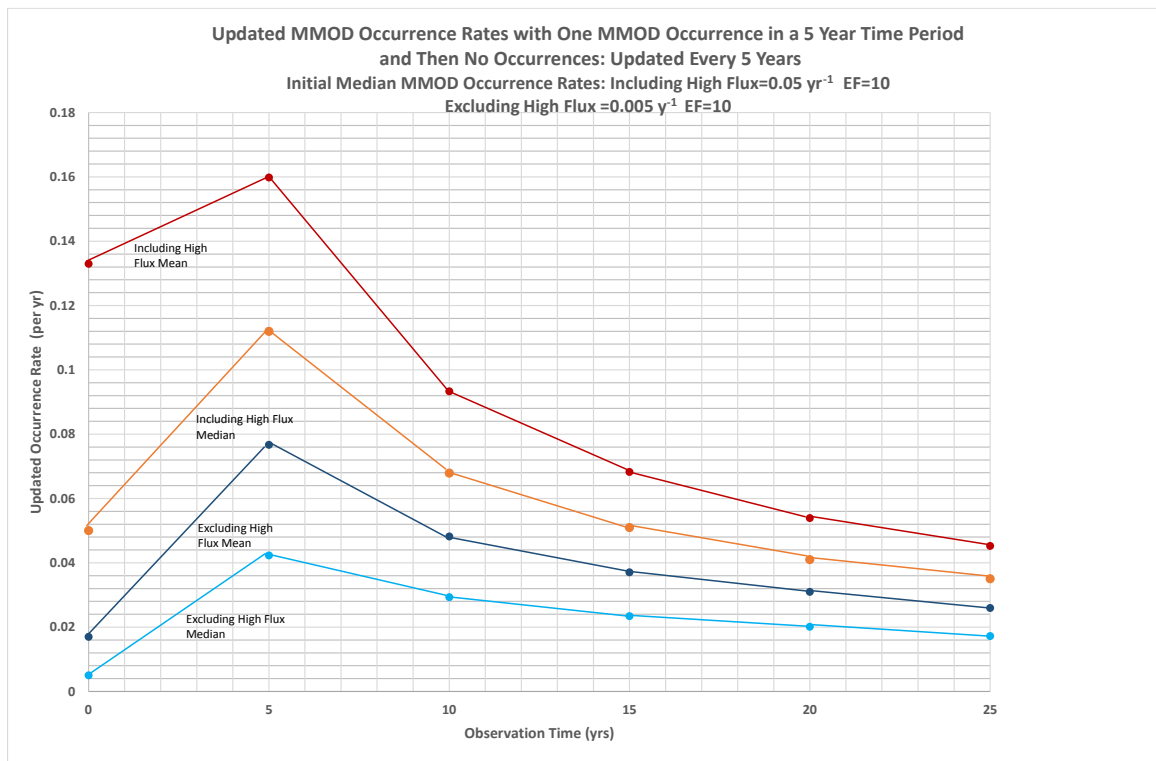


Figure 15. Updated Mean and Median MMOD Occurrence Rates for Two Alternative Models for Observed Data of One Occurrence in 5 Years then No Occurrences

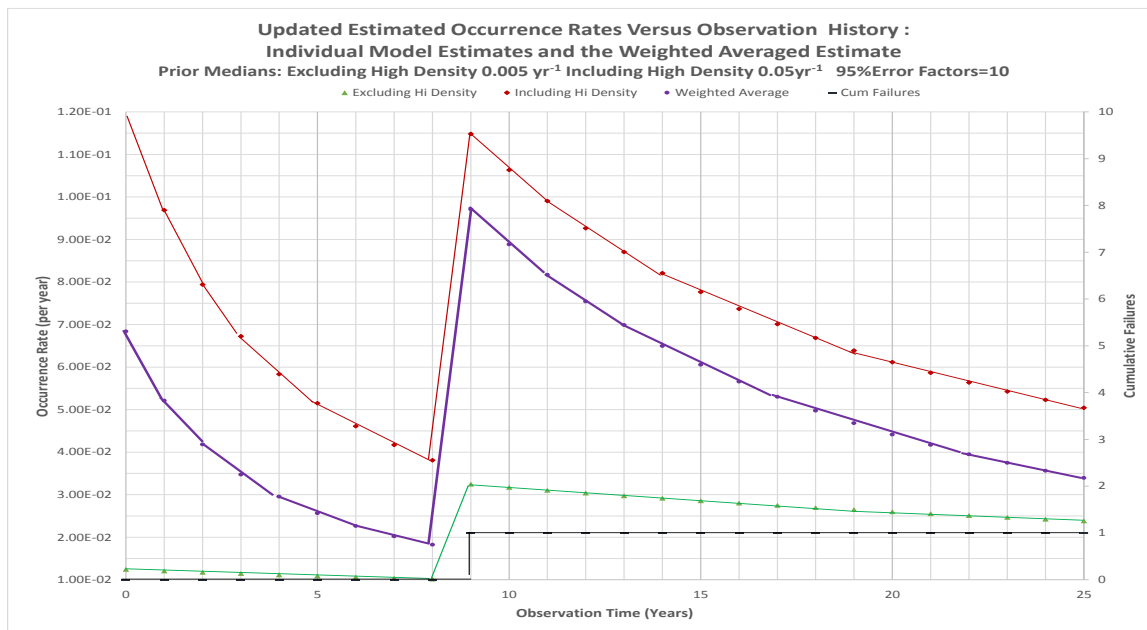


Figure 16. Yearly Updated Median MMOD Occurrence Rate for Two Alternative Models for Observed Data of One Occurrence between 8 and 9 Years then No Occurrences

References

1. NASA/SP-2009-569, *Bayesian Inference for NASA Probabilistic Risk and Reliability Analysis*, 2009
2. Sharon Bertsch McGrayne, *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy*, Yale University Press, 2012
3. International Society for Bayesian Analysis, *What is Bayesian Analysis*, <http://bayesian.org/>
4. Gregory, P., *Bayesian Logical Data Analysis for the Physical Sciences*, Cambridge University Press, 2005
5. Joseph K. Blitzstein and Jessica Hwang, *Introduction to Probability*, Chapman & Hall/CRC Texts in Statistical Science, 2014
6. MD Anderson Center, Parameter Solver, https://biostatistics.mdanderson.org/SoftwareDownload/SingleSoftware.aspx?Software_Id=6
7. Catherine Forbes, Merran Evans, Nicholas Hastings, and Brian Peacock, *Statistical Distributions*, Wiley, 2010
8. Reuven Y. Rubinstein and Dirk P. Kroese, *Simulation and the Monte Carlo Method*, Wiley, 2007
9. Andre I. Khuri and John A. Cornell, *Response Surfaces: Designs and Analyses*, Marcel Dekker, 1996
10. Norman Fenton and Martin Neil, *Risk Assessment and Decision Analysis with Bayesian Networks*, CRC Press, 2012
11. Finn V. Jensen and Thomas Nielsen, *Bayesian Networks and Decision Graphs*, Springer-Verlag, 2002
12. Geoffrey McLachlan and David Peel, *Finite Mixture Models*, Wiley, 2000
13. Kerrie Mengersen, Christian Robert, and Mike Titterton, *Mixtures: Estimation and Applications*, Wiley, 2011
14. NASA-HDBK-7009, *NASA Handbook for Models and Simulations: An Implementation Guide for NASA-STD-7009*, 2013
15. Robert J. Serfling, *Approximation Theorems of Mathematical Statistics*, Wiley, 2001
16. Christopher Small, *Expansions and Asymptotics for Statistics*, Chapman & Hall/CRC, 2010
17. Andrea Saltelli, Stefano Tarantola, Francesca Campolongo, and Marco Ratto, *Sensitivity Analysis in Practice*, Wiley, 2004
18. Bradley Carlin and Thomas A. Louis, *Bayesian Methods for Data Analysis*, Chapman & Hall/CRC, 2008
19. Ronald Christensen, Wesley Johnson, Adam Branscum, and Timothy Hansen, *Bayesian Ideas and Data Analysis: An Introduction for Scientists and Statisticians*, Chapman & Hall/CRC, 2010
20. George A. Milliken, *Analysis of Messy Data*, Chapman & Hall/CRC, 2004
21. Michael Berthold and David Hand, *Intelligent Data Analysis*, Springer-Verlag, 2007
22. James Berger, *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag, 2010
23. Jim Q. Smith, *Bayesian Decision Analysis: Principles and Practice*, Cambridge Press, 2010

Appendices

Examples of WINBUGS Scripts for a Poisson Likelihood and Lognormal Uncertainties

```
# POISSON WITH LOG NORMAL PRIOR
```

```
model{
  for(i in 1:5){
    x[i]~dpois(mean.poisson[i])
    mean.poisson[i]<-lambda[i]*time.yr[i]
    lambda[i]~dlnorm(mu,tau)
    prob[i]<-pow(mean.poisson[i],x[i])*exp(-
    mean.poisson[i])/exp(logfact(x[i]))
  }
}
```

```
tau<-1/pow(log(prior.EF)/1.645,2)
mu<-log(prior.median)
}
```

```
Data
```

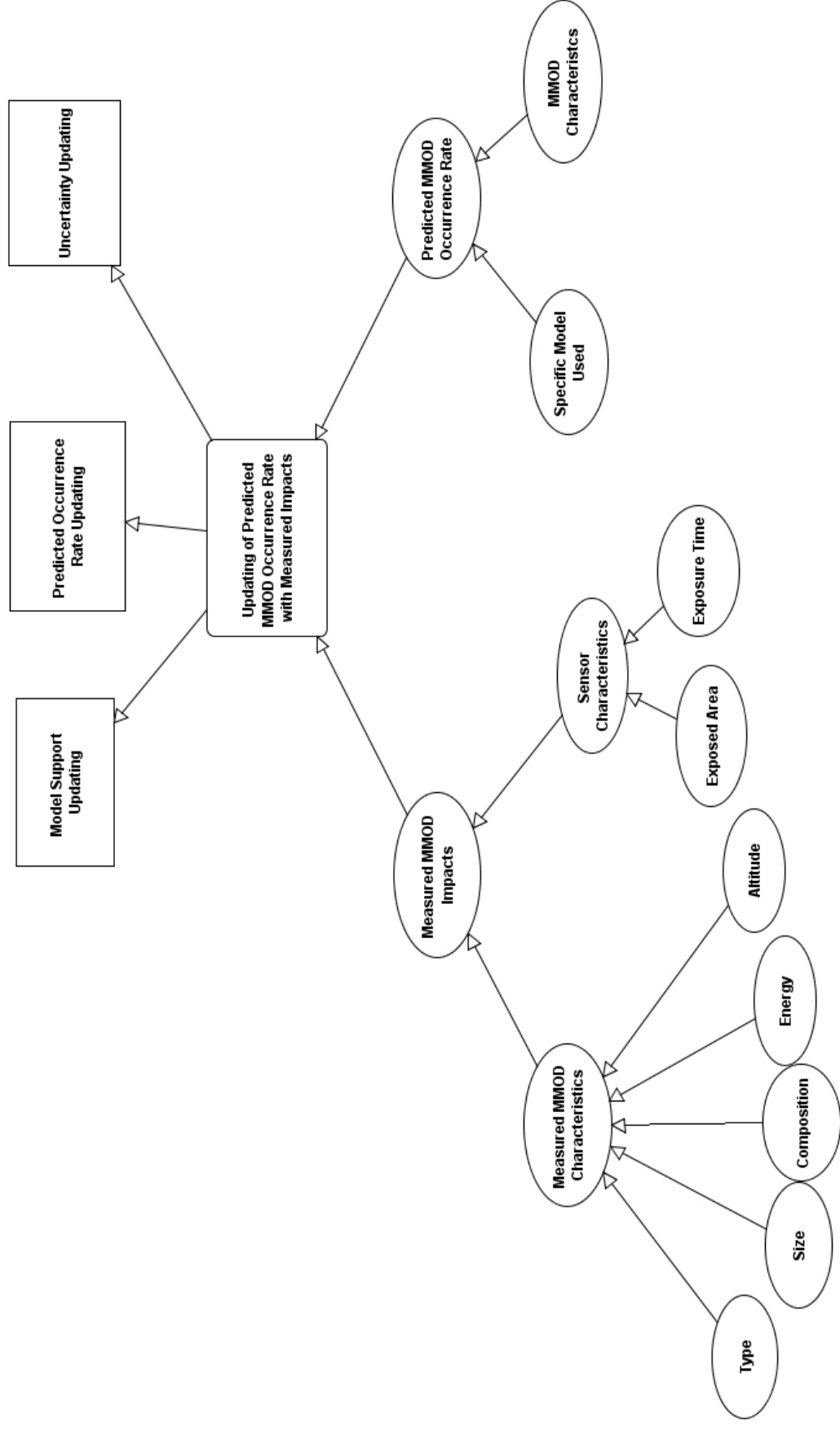
```
list(x=c(0,0,0,0,0), time.yr=c(5,10,15,20,25), prior.median=.05,
prior.EF=10)
```

```
# POISSON WITH LOG NORMAL PRIOR WITH WEIGHTED AVERAGE
```

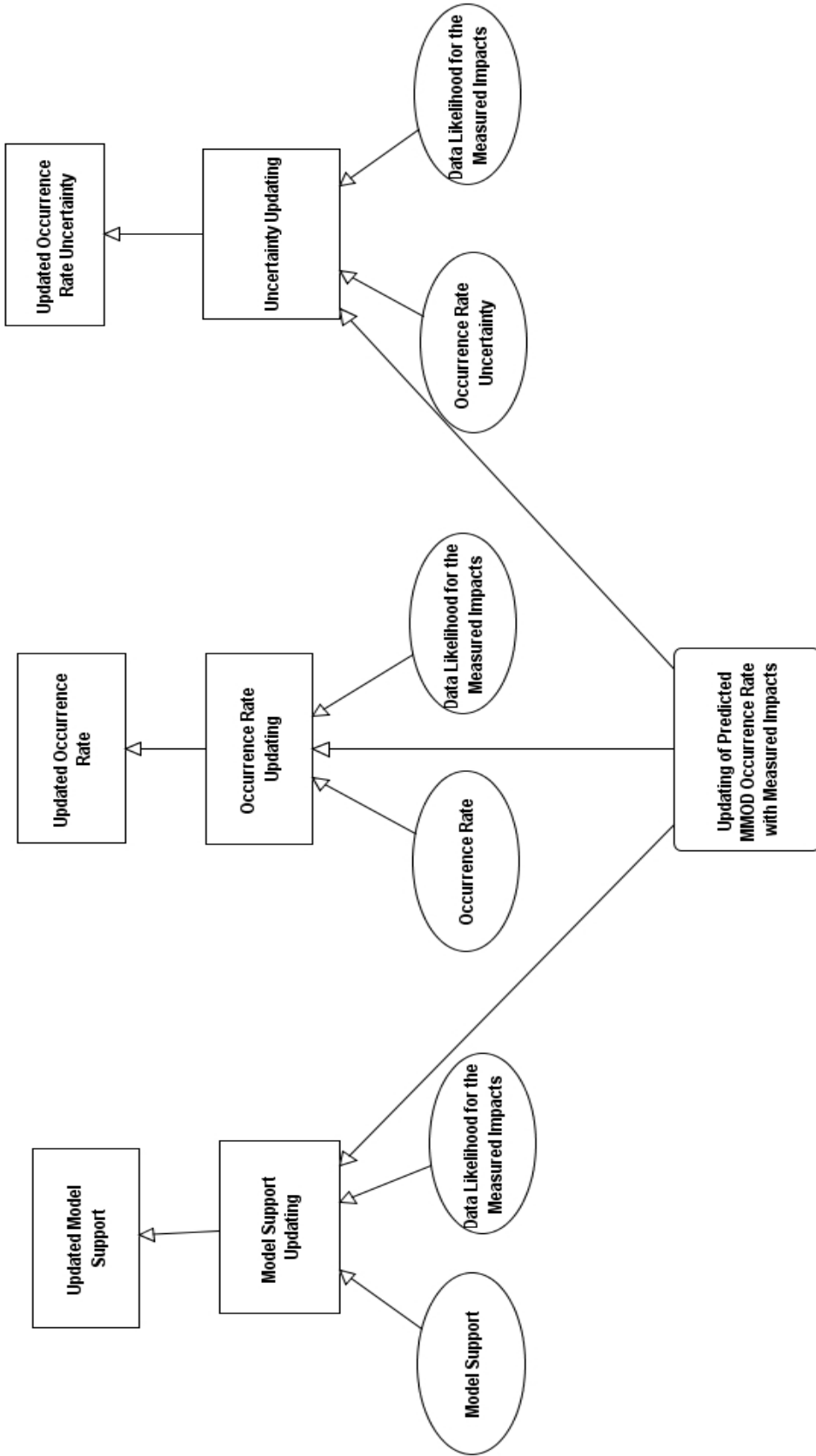
```
model{
  for(i in 1:5){for(j in 1:2){
    x[i,j]~dpois(mean.poisson[i,j])
    mean.poisson[i,j]<-lambda[i,j]*time.yr[i,j]
    lambda[i,j]~dlnorm(mu[j],tau[j])
    prob[i,j]<-pow(mean.poisson[i,j],x[i,j])*exp(-mean.poisson[i,j])/exp(logfact(x[i,j]))}
    wt.avg[i]<-(p1[i]/(p1[i]+p2[i]))*lambda[i,1]+(p2[i]/(p1[i]+p2[i]))*lambda[i,2]}
  for(k in 1:2){
    tau[k]<-1/pow(log(prior.EF[k]/1.645),2)
    mu[k]<-log(prior.median[k])
  }
}
```

```
Data
```

```
list(x=structure(.Data=c(0,0,0,0,0,0,0,0,0,0), .Dim=c(5,2)),
time.yr=structure(.Data=c(5,5,10,10,15,15,20,20,25,25), .Dim=c(5,2)),
prior.median=c(.05,.005), prior.EF=c(10,10), p1=c(.7958,.7245,.681,.647,.6224),
p2=c(.9521,.9222,.8957,.877,.8616))
```



Flow Chart for Designing Experiments to Update MMOD Risk Estimates with Observed Data



Analysis Flow Chart for Designing Experiments to Update MMOD Risk Estimates with Observed Data