5th Annual Symposium, Mary Kay O'Connor Process Safety Center
"Beyond Regulatory Compliance: Making Safety Second Nature"
Reed Arena, Texas A&M University, College Station, Texas
October 29-30, 2002

Probabilistic Fault Tree Analysis

Sanjeev Mohindra
TIAX LLC
20 Acorn Park
Cambridge MA 02140 USA
Phone: (617) 498-5831
Email: mohindra.s@tiax.biz

Introduction

Any system failure model contains at least two sources of uncertainty; modeling uncertainty, and parametric uncertainty. One of the main causes of parametric uncertainty is the underlying data evaluation uncertainty. Very often failure data for events is not available and engineering estimates are used. The engineering estimate (expert estimates) of component reliability parameters often results in extreme data uncertainty. While there exist importance measures to evaluate the individual contribution of component failures to the overall system failure, they fail to provide information about the overall uncertainty at the system level.

Uncertainty quantification at the system level must be part of the final decision–making process. There is often significant uncertainty in the scenario frequency, and thus in the overall risk curve and loss profile. Knowledge of the uncertainty in the overall risk curve is import to the decision-maker; he/she can make more informed decisions. It is also important for the risk analyst; more effort can be made in reducing the uncertainty of key events.

Basic Fault Tree Analysis

Fault Tree analysis (FTA) is a system level deductive method for determining the various combinations of hardware failures, software failures, and human errors that could result in the occurrence of an undesired event (referred to as top event). The main purpose of fault tree analysis is to evaluate the probability or the frequency of the top event using quantitative information about the causal events (referred to as basic events). FTA can provide useful information concerning the likelihood of a failure and the means by which such a failure could occur. Efforts to improve system safety and reliability can be focused and refined using the results of the FTA.

A basic event represents a simple failure or fault. It may be a hardware failure, a human error, or an adverse environment condition. Hardware failures are usually expressed in terms of a specific component and a failure mode, such as "Service Water Pump P-123 fails to start on demand." Human errors can be failure to carry out a desired task (failure to open a valve), failure to perform a specific recovery action (failure to start a backup system), or execution of a wrong action that has adverse effects on the fault tree top event. An adverse environment condition is not necessarily a failure but in combination with other events can lead to failure. For example, the temperature being below freezing is an adverse condition necessary for the failure of flow reduction due to a frozen pipe.

Basic events are assumed to be independent of each other. This means that the occurrence of one basic event does not influence the probability of occurrence of any other basic event. For example, suppose that there are two pumps, and the failure of either to start on demand is a basic event. Independence of the basic events says that if one pump fails to start on demand, this does not alter the probability that the second pump will fail to start.

A common cause event, such as "two pumps fail to start because of loss of power" must be modeled as its own basic event, and be assigned its own failure probability or failure rate. This event is then regarded as statistically independent of all other basic events.

A Fault Tree Example

To guide the discussion through the paper, we will make use of an example fault tree. This example, while not "industrial strength", provides all the features that we need for our discussions. The frequencies, and probabilities used in the fault tree are merely representative values used for the purpose of the example.

The system [1] in Figure 1 is designed to 1) decrease the temperature of the hot gas by a water quench, 2) saturate the gas with water vapor, and 3) remove solid particles entrained in the gas. The hot tail gas is first cooled by contacting it with water supplied by the Feedwater pump D. Water from the bottom of the scrubber is either recirculated by pumps E or F, or removed as a purge stream. Mesh filter pad G removes the particulates from the gases that flow to an absorber. A simplified fault tree is shown in Figure 2.

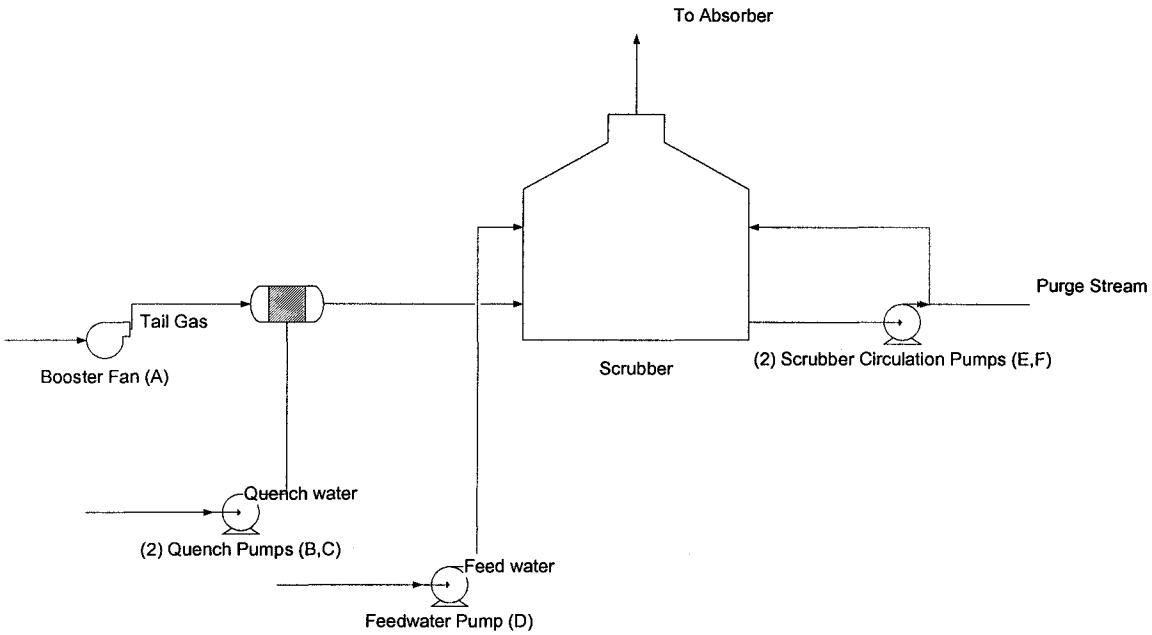


Figure 1: Example Problem

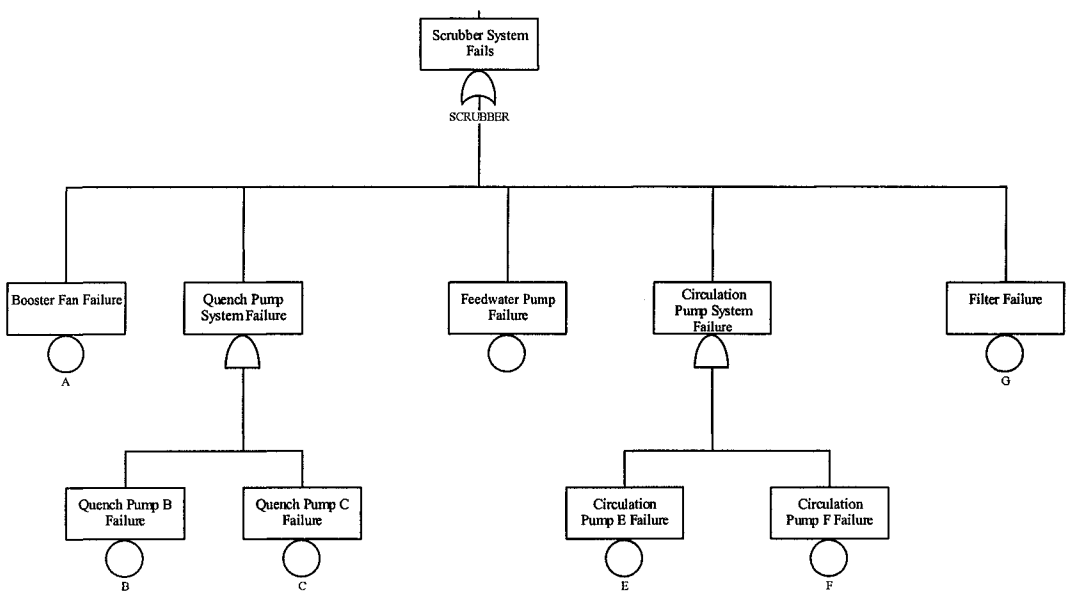


Figure 2: Fault tree for the process shown in Figure 1

Quantifying Fault Trees

Quantifying fault trees involves the following: (1) quantification of basic events, (2) decomposition of the fault tree into cut sets, (3) calculation of individual cut set probabilities, and (4) quantification of the top event based on the individual cut set probabilities.

1. *Quantification of the basic events*

The basic event probabilities or frequencies can be estimated from available literature data, and/or plant specific data. Some common sources of information are provided in the reference section.

2. *Decomposition of the fault tree into cut sets*

The logic for each gate starting with the top event is recursively replaced with its inputs until the resulting logic is expressed in terms of basic events only. This results in a list of cut sets of the fault tree. A cut set identifies a set of events that will cause the top event to occur. The list of cut sets identifies all such logical combinations of events. Once the minimal cut sets have been determined; individual cut set probabilities can be calculated.

3. *Calculation of individual cut set probabilities (or frequencies)*

The individual cut set probabilities are determined by multiplying the probabilities of the applicable basic events. This assumes statistical independence of the basic events.

4. *Calculation of top event probabilities (or frequencies)*

The exact top event probability of the union of the cut sets can be found, in principle, by combining the probabilities of each cut set. For large fault trees, this exact method can be too time consuming. Two approximations -- the rare event approximation and the minimal cut set upper bound are often used to reduce the computational burden.

Rare Event Approximation

A common approach to calculate the probability for a top event is to add together the probabilities for the individual cut sets. The rare event approximation is good when the cut set probabilities are small.

$$P = \sum_{i=0}^{m} C_i$$

where,

P = rare event approximation for system unavailability,
C_i = probability of the i^{th} cut set, and
m = number of minimal cut sets in the fault tree.

Minimal Cut Set Upper Bound

The minimal cut set upper bound is calculated using the following equation

$$1 - S = \prod_{i=1}^{m} (1 - C_i)$$

where

S = minimal cut set upper bound for the system unavailability,
C_i = probability of the i^{th} cut set, and
m = number of minimal cut sets in the fault tree.

Using fault trees to make decisions
It is often difficult to make decisions based just on the top event probability or frequency without understanding the important contributory events, or the uncertainty in the estimate. Several measures exist for determining the importance of individual basic events on the top events. These measures are detailed in the next section.

Measuring the Importance
There are several different basic event importance measures. These importance measures are calculated for each basic event in a fault tree. The main importance measures are:

- Fussell-Vesely importance, an indication of the percentage of the minimal cut set upper bound contributed by the cut sets containing the basic event
- Risk reduction , an indication of how much the minimal cut set upper bound would decrease if the basic event never occurred (typically, if the corresponding component never failed)
- Risk increase , an indication of how much the minimal cut set upper bound would go up if the basic event always occurred (typically, if the corresponding component always failed)
- Structural importance, the number of cut sets that contain the basic event.

The importance measures can be defined as ratios or as differences. The ratio importance measures are dimensionless, and are appropriate for purely relative evaluations. The difference definitions account for the actual risk levels that exist and are more appropriate when actual risk levels are of concern. Let us define the following:

F(x) = minimal cut set upper bound probability (frequency) evaluated with the basic event probability(frequency) at its mean value.

F(0) = minimal cut set upper bound probability (frequency) evaluated with the basic event probability set to 0.0.

F(1) = minimal cut set upper bound probability (frequency) evaluated with the basic event failure probability set to 1.0.

Fussell-Vesely Importance

The Fussell-Vesely importance is an indication of the fraction of the minimal cut set upper bound (or frequency) that involves the cut sets containing the basic event of concern. It is calculated by finding the minimal cut set upper bound of those cut sets containing the basic event of concern and dividing it by the minimal cut set upper bound of the top event. This calculation can be performed by determining the minimal cut set upper bound with the basic event failure probability at its mean value and again with the basic event failure probability set to zero. The Fussell-Vesely importance FV can then be calculated as:

$$FV = [F(x) - F(0)]/F(x)$$

Risk Reduction

The risk reduction importance measure is an indication of how much the results would be reduced if the specific event probability equaled zero, normally corresponding to a totally reliable piece of equipment. The risk reduction ratio is determined by evaluating the fault tree minimal cut set upper bound (or the frequency) with the basic event probability set to its true value and dividing it by the minimal cut set upper bound (frequency) calculated with the basic event probability set to zero. The risk reduction ratio RRR is:

$$RRR = F(x)/F(0)$$

The risk reduction difference indicates the same characteristic as the risk reduction ratio, but it reflects the actual minimal cut set upper bound levels instead of a ratio. This is the amount by which the failure probability or sequence frequency would be reduced if the basic event never failed.

The risk reduction difference (RRD) is calculated by taking the difference between the mean value and the function evaluated at 0. The risk reduction difference RRD is:

$$RRD = F(x) - F(0)$$

Risk Increase

The risk increase ratio is an indication of how much the top event probability (frequency) would go up if the specific event had probability equal to 1.0, normally corresponding to totally unreliable equipment. The risk increase ratio is determined by evaluating the minimal cut set upper bound (sequence frequency) with the basic event probability set to 1.0 and dividing it by the minimal cut set upper bound evaluated with the basic event probability set to its true value. The risk increase ratio RIR is:

$$RIR = F(1)/F(x) .$$

The risk increase difference RID is calculated by taking the difference between the function evaluated at 1.0 and the nominal value. The risk increase difference RID is:

$RID = F(1) - F(x)$.

Birnbaum Importance

The Birnbaum importance measure is similar to the FussellVesely importance measure except that it deals with differences instead of ratios. The Birnbaum importance is an indication of the sensitivity of the minimal cut set upper bound (or frequency) with respect to the basic event of concern. It is calculated by determining the minimal cut set upper bound (or frequency) with the basic event probability of concern set to 1.0 and again with the basic event probability set to 0.0. The difference between these two values is the Birnbaum importance. In equation form, the Birnbaum importance B is
$B = F(1) - F(0)$.

Measuring the uncertainty

The uncertainty analysis allows the user to calculate the uncertainty in the top event probability (or frequency) resulting from uncertainties in the basic event probabilities (or frequencies). To use this option, the user must have the component reliability information and distribution data.

Before performing the uncertainty analysis, the top event is often expressed in terms of minimal cut sets. These cut sets depend on many basic events, each of which has a probability described in terms of some parameter(s). Suppose that a basic event probability is **p**. The value of **p** for each basic event is not known exactly, but is estimated based on data or on expert opinion. The uncertainty in **p** is quantified by a probability distribution: the mean of the distribution is the best estimate of **p**, and the dispersion of the distribution measures the uncertainty in **p**, with a large or small dispersion reflecting large or small uncertainty, respectively, in the true value of **p**. This distribution is the uncertainty distribution of **p**.

For all the basic events, random samples of the probability are taken based on the uncertainty distributions. These sampled probability values are then used to calculate the probability of the top event. These sampling and top event calculations are repeated many times, and the uncertainty distribution for the probability of the top event is thus found empirically. The mean of the distribution is the best estimate of the probability of the top event, and the dispersion quantifies the uncertainty in this probability. The term Monte Carlo is used to describe this analysis by repeated random sampling.

Overview of Simple Monte Carlo Sampling

The Monte Carlo approach is the most fundamental approach to uncertainty analysis. Simple Monte Carlo simulation consists of making repeated quantifications of the top event value using values selected at random from the uncertainty distributions of the basic events. In every iteration of the Monte Carlo run, each basic event uncertainty distribution is sampled using a random number generator to select the failure probability

of the basic event. The top event probability or accident sequence frequency is calculated, and the uncertainty distribution is also generated.

To illustrate the Monte Carlo technique, consider a system with two components in series. Let A denote failure of the first component and B failure of the second. The cut sets for the system are A and B, so the equation for the top event (system) is

$$S = A + B$$

Let A and B have mean failure probabilities of 0.001 and 0.005, respectively. Also assume that the uncertainty distribution for A is uniform from 0 to 0.002 and the distribution for B is normal with standard deviation of 0.001.

The point estimate for S is 0.006. The probability distributions for A, B, and S are shown in Figure 3.

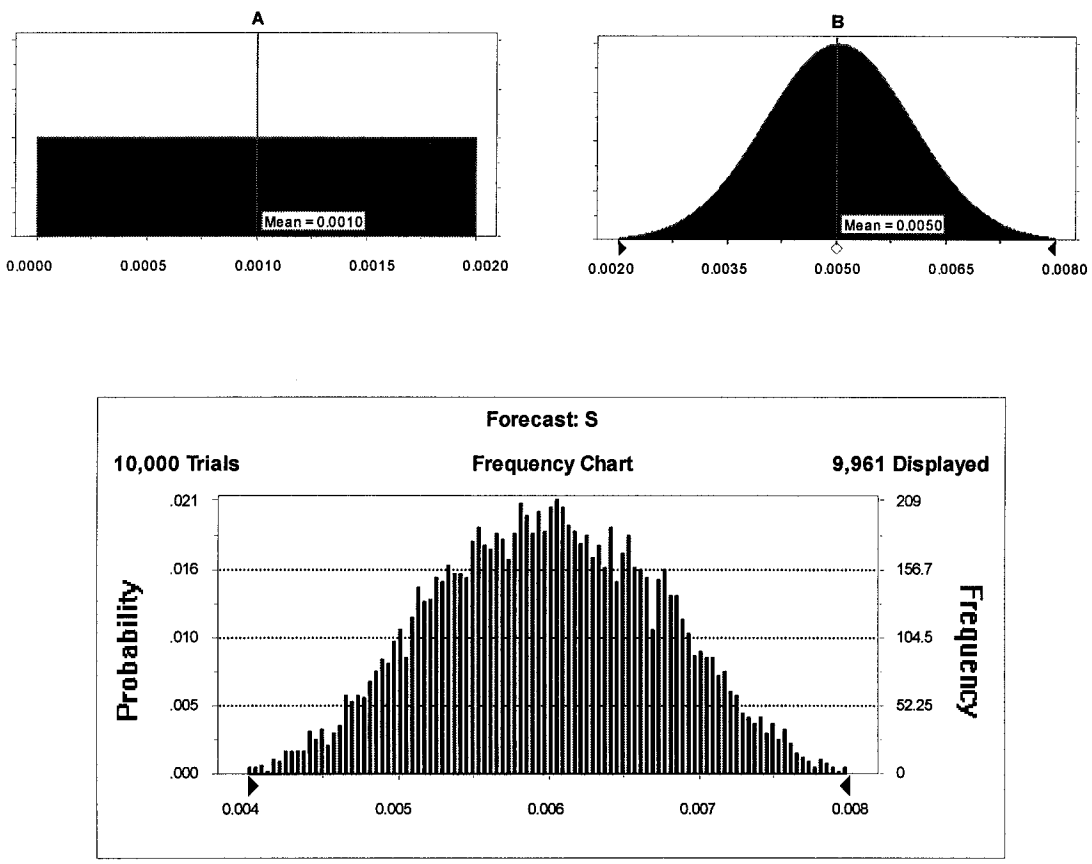


Figure 3: Input and output distributions for S=A+B

In addition to the graphical representation, we also get a detailed statistical picture, as shown in Table 1

Table 1: Statistical parameters for S=A+B

Statistic	Value
Trials	10000
Mean	0.005996486
Median	0.005995968
Standard Deviation	0.000757095
Variance	5.73192E-07
Skewness	0.00
Kurtosis	2.58

Input Data Distributions

For uncertainty analysis, the basic event data needs to be provided as a probability distribution. The most common distribution used for uncertainty analysis is the lognormal distribution. Other distributions that are sometimes used are normal, beta, gamma, chi-squared, exponential, and uniform distributions.

Most distributions can be defined with two statistical parameters, although some take more. The first parameter is the mean failure probability and the second parameter is specific to the particular uncertainty distribution. Table 2 summarizes this information for some common distributions.

Table 2: Common data distributions

Distribution	Parameter
lognormal	95% error factor
normal	Standard deviation
beta	b in beta(a, b)
gamma	r in gamma(r)
chi-squared	Degrees of freedom
Exponential	-
Uniform	Upper end point

Applying Monte Carlo method to our sample example

Let us now apply Monte Carlo analysis techniques to the fault tree shown in Figure 2. We will assume that all the basic events are probabilities (or have been reduced to probabilities). The distributions for the basic events are shown in Figure 4. The results of Monte Carlo simulation are shown in Figure 5 and Figure 6. Note that the top event estimate is calculated using the minimal cut set upper bound assumption for simplicity.

As can be seen from the results, we get the probability of the top event and a visual representation of the uncertainty. The tornado chart provides the relative significance of

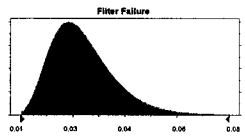
the individual basic events. This can provide confidence (or lack thereof) in the analysis and point out deficiencies in the data. While not apparent in this simple example, the top event distribution may have several peaks, and a visual representation is needed to get a feel for the results.

Basic Event Models

Filter Failure

Lognormal distribution with parameters:
Mean 0.03
Standard Dev. 0.01

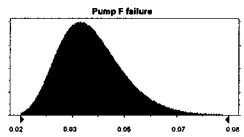
Selected range is from 0.00 to +Infinity



Pump F failure

Lognormal distribution with parameters:
Mean 0.04
Standard Dev. 0.01

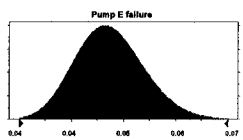
Selected range is from 0.00 to +Infinity



Pump E failure

Lognormal distribution with parameters:
Mean 0.05
Standard Dev. 0.01

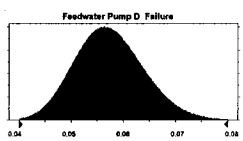
Selected range is from 0.00 to +Infinity



Feedwater Pump D Failure

Lognormal distribution with parameters:
Mean 0.06
Standard Dev. 0.01

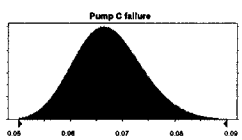
Selected range is from 0.00 to +Infinity



Pump C failure

Lognormal distribution with parameters:
Mean 0.07
Standard Dev. 0.01

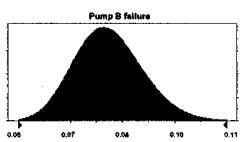
Selected range is from 0.00 to +Infinity



Pump B failure

Lognormal distribution with parameters:
Mean 0.08
Standard Dev. 0.01

Selected range is from 0.00 to +Infinity



Booster Fan Failure

Normal distribution with parameters:
Mean 0.09
Standard Dev. 0.00

Selected range is from -Infinity to +Infinity

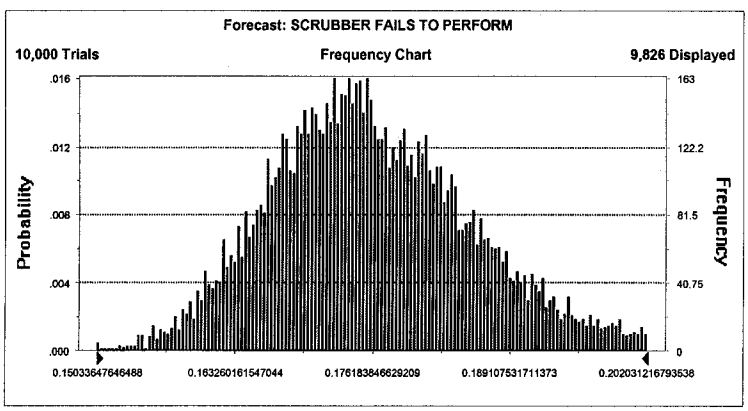
Figure 4: Basic event models for example

Forecast: SCRUBBER FAILS TO PERFORM

Summary:

Display Range is from 0.15033647646488 to 0.202031216793538
Entire Range is from 0.145040841869696 to 0.232119569296944
After 10,000 Trials, the Std. Error of the Mean is 0.00010089543440364

Statistics:	Value
Trials	10000
Mean	0.176561115
Median	0.175482762
Mode	---
Standard Deviation	0.010089543
Variance	0.000101799
Skewness	0.67
Kurtosis	3.99
Coeff. of Variability	0.06
Range Minimum	0.145040842
Range Maximum	0.232119569
Range Width	0.087078727
Mean Std. Error	0.000100895



Forecast: SCRUBBER FAILS TO PERFORM (cont'd)

Percentiles:

Percentile	Value
0%	0.145040842
10%	0.164717042
20%	0.168140393
30%	0.17083702
40%	0.173226312
50%	0.175482762
60%	0.178016594
70%	0.180943869
80%	0.184283921
90%	0.189560178
100%	0.232119569

Figure 5: Top event probability for Scrubber fails to perform

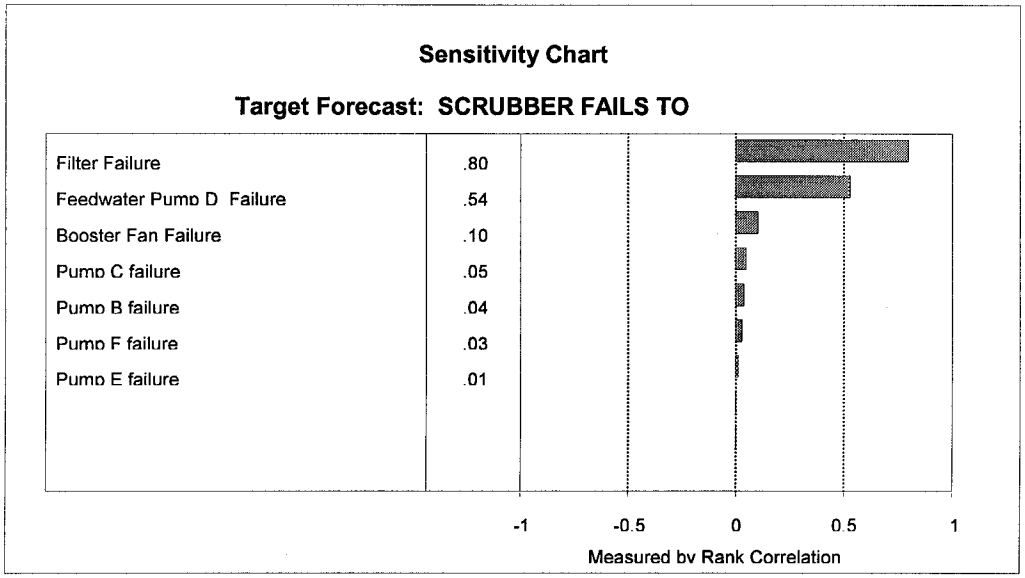


Figure 6: Sensitivity of top event to basic events

Conclusion

Uncertainty quantification at the system level must be part of the final decision–making process. Knowledge of the uncertainty in the overall risk curve is import to the decision-maker; he/she can make more informed decisions. It is also important for the risk analyst; he/she can allocate more effort to better quantify basic events that contribute more to the top event.

References

1. Kumamoto, H. and E.J. Henley, *Probabilistic Risk-Analysis and Management for Engineers and Scientists* (2nd edition), IEEE Press (1996).
2. *Offshore Reliability Data Handbook* 3rd Edition, 1997.
3. Lees, F.P. *Loss Prevention in the Process Industries*, Vols 1- 3, Butterworth, 1995
4. Cluley, J.C. *Reliability in Instrumentation and Control,*, Butterworth-Heinemann, Boston, 1993.
5. IEEE Std. 500. *Guide to the Collection and Representation of Electrical, Electronic, Sensing Component, and Mechanical Equipment Reliability Data for Nuclear Generating Stations*. IEEE, New York, 1984.