IECM Technical Documentation:

Uncertainty Analysis

January 2019

IECM Technical Documentation:

Uncertainty Analysis

Prepared by:

The Integrated Environmental Control Model Team
Department of Engineering and Public Policy
Carnegie Mellon University
Pittsburgh, PA 15213
www.iecm-online.com

For
U.S. Department of Energy
National Energy Technology Laboratory
P.O. Box 880

Compiled in January 2019

Table of Contents

List of Figures

Acknowledgements

This documentation is a compilation of the following textbook, report, and paper, where were referred to for developing the uncertainty engine in the Integrated Environmental Control Model (IECM):

- Diwekar, U. M., & Rubin, E. S. (1991). Stochastic modeling of chemical processes. *Computers & Chemical Engineering*, *15*(2), 105-114.
- Morgan, M. G., Henrion, M., & Small, M. (1992). Uncertainty: a Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. Cambridge university press.
- Swiler, L.P., & Wyss, G.D. (2004). A User's Guide to Sandia's Latin Hypercube Sampling Software: LHS UNIX Library/Standalone Version; No. SAND2004-2439, Sandia National Laboratories, Albuquerque, NM.

Uncertainty Analysis

Introduction

Sensitivity analysis is often conducted to examine the effects of variable parameters. However, it typically examines one parameter or two at a time with all other parameters held constant. In some cases, however, several parameters may be varied simultaneously. Thus, interactions among several uncertain parameters may be overlooked. Furthermore, sensitivity analysis cannot provide information on the likelihood of specific outcomes. The robust systems analysis requires both deterministic and stochastic modeling capabilities. The ability to characterize uncertainties explicitly is a feature unique to the Integrated Environmental Control Model (IECM). Numerous input parameters can be assigned probability distributions. When input parameters are uncertain, an uncertainty distribution of results is returned. Such result distributions give the likelihood of a particular value, in contrast to conventional single-value estimates. This document gives a brief introduction to probability distribution functions, characterization parameters, and sampling methods used in the IECM.

Characterization of Probability Distributions

The uncertainty engine in the IECM referred to Chapter 5 of the uncertainty textbook by Morgan *et al.* (1990). This section is a simplified version of relevant contents presented in that chapter.

Probability Distributions

The likelihood of a random variable having different possible values is represented by a probability distribution. Suppose that X is a random variable and *x* is a particular possible value that X might have, the probability distribution for X may be described by its *cumulative distribution function* (CDF):

$$F(X) \equiv P[X \leq x]$$

This function gives the cumulative probability that X will be less than or equal each possible value *x*. When *x* increases from its minimum to maximum values, F(*x*) increases monotonically and goes from 0 to 1.

The probability distribution for X may also be described by its *probability density function* (PDF):

$$f(X) = \frac{dF(X)}{dX}$$

The PDF is the derivative of the CDF and represents the density of probability. The values of X that have maximum probability density are called *modes*.

Fractiles

The *p* fractile or quantile of a distribution represents a probability of *p* that the actual value of a random variable will be less than that value:

$$P\left[X \leq X_p\right] \equiv p$$

The *median* of a random variable is the 0.5 fractile of a distribution. The 0.25 and 0.75 fractiles are often called quartiles of the distribution. The range of $[X_{0.25}, X_{0.75}]$ is called the *interquartile range* of a distribution, which measures the distribution dispersion. If the probability is expressed in percent, the fractile value is also called a *percentile*.

Moments

Moments are often used to characterize probability distribution functions. The *mean* or *expected value* of the distribution is the first moment and is defined for a continuous and discrete variable, respectively as:

$$\mu = \int_X xf(x)dx$$

$$\mu = \sum_{i=1}^{n} x_i p(x_i)$$

Central moments are defined as the expectation of the nth power of the difference between X and its expected value:

$$\mu_n = \int_X (x-\mu)^n f(x)dx$$

$$\mu_n = \sum_X (x-\mu)^n \; p(x)$$

The *variance* (σ^2) is the second central moment, and its square root is the *standard deviation*. They indicate the degree of distribution spread. The coefficient of variation characterizes the shape of a distribution and is defined as the ratio of the standard deviation to the mean.

The *coefficient of skewness* is the third central moment that characterizes and classifies distributions and is defined as:

$$\gamma_1 = \frac{\mu_3}{\sigma^3}$$

The *coefficient of kurtosis* is the fourth central moment that characterizes and classifies distributions and is defined as:

$$\gamma_2 = \frac{\mu_4}{\sigma^4}$$

The skewness coefficient is zero for a symmetric distribution. The kurtosis coefficient indicates the degree to the which the distribution is flat as opposed to the one having a high central peak.

Probability Distribution Functions

This section briefly summarizes the functions and feature parameters of probability distributions. Figure 1 shows several examples of probability distribution functions available in the IECM library. In addition to the common probability distributions, a user-specified distribution option is also available in the IECM. More details about probability distributions are discussed in Chapter 5 of the uncertainty textbook by Morgan *et al.* (1990).

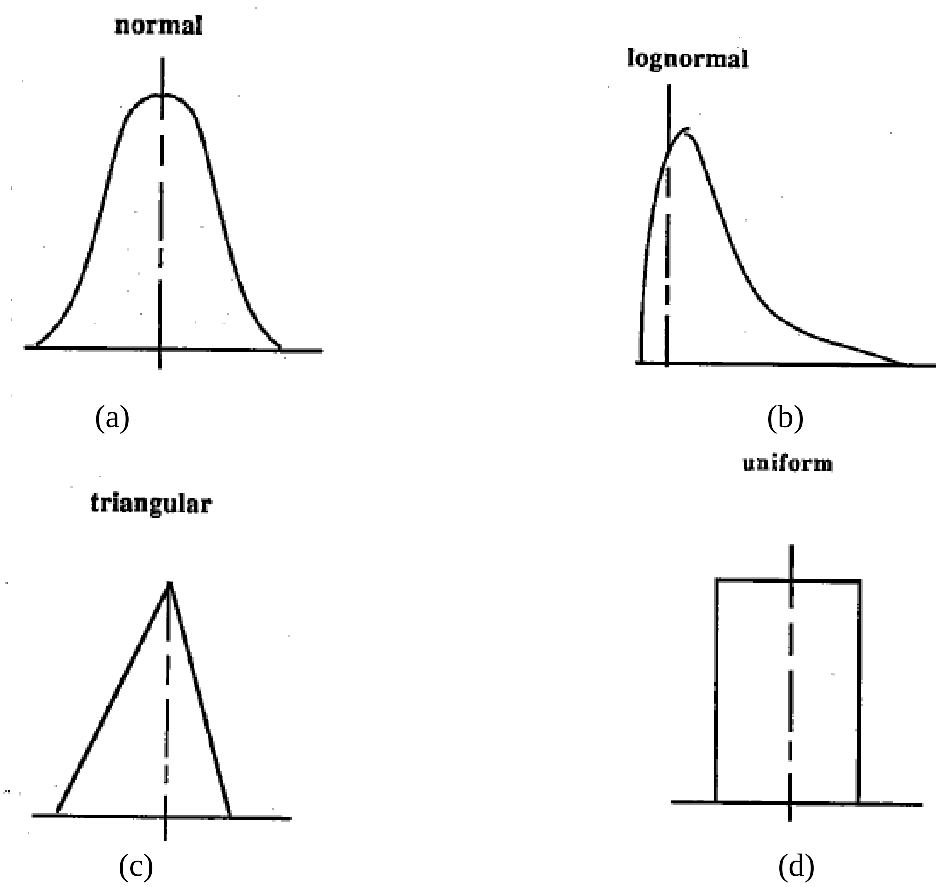


Figure 1. Probability Distribution Examples

(Source of the figures: Diwekar and Rubin 1991)

A probability distribution shows the range of value a variable could take and the likelihood of occurrence of each value within the range (Diwekar and Rubin 1991). A normal distribution represents a symmetric probability of a parameter value being more or less than the mean value. In contrast, lognormal and triangular distributions are skewed so that there is a higher probability of values standing on one side of the median than the other. A uniform distribution represents an equal likelihood of a value within a range. They are summarized as follows (Morgan *et al.* 1990):

Normal Distribution:

PDF: $f(x) = \frac{1}{(2\pi)^{1/2}\sigma} exp(-\frac{(x-\mu)^2}{2\sigma^2}), -\infty \le x \le \infty$

CDF: no closed-form representation

Moments: $\mu, \sigma, \gamma_1 = 0, \gamma_2 = 3$

Parameter Estimation: $\hat{\mu} = \bar{X}, \hat{\sigma} = S = \left(\frac{1}{m-1}\sum_{i=1}^{m}(x_i - \bar{x})^2\right)^{1/2}$

Lognormal Distribution:

PDF: $f(x) = \frac{1}{\emptyset x(2\pi)^{1/2}} exp(-\frac{(lnx-\xi)^2}{2\emptyset^2}), 0 \le x \le \infty$

Where ξ and $\emptyset$ are the paremters of the distribution.

CDF: computed from the normal distribution of lnx.

Moments: with $\omega = exp(\phi^2), \mu = exp(\xi + \phi^2/2), \sigma^2 = \omega(\omega - 1)exp(2\xi)$

$$\gamma_1 = (\omega - 1)^{1/2}(\omega + 2), \gamma_2 = \omega^4 + 2\omega^3 + 2\omega^2 - 3$$

Parameter Estimation: $\hat{\xi} = \frac{1}{n}\sum_{i=1}^{n} lnx_i, \hat{\phi} = \left(\frac{1}{n}\sum_{i=1}^{n}(lnx_i - \hat{\xi})^2\right)^{1/2}$

Triangular Distribution:

PDF: $f(x) = \frac{b-|x-a|}{b^2}, a - b \le x \le a + b$

Where a and b are the paremters of the distribution.

CDF: $F(x) = \frac{1}{2} + \frac{1}{b^2}\left(\frac{1}{2}(x^2 - a^2) + (b - a)(x - a)\right); a - b \le x \le a$

$$= 1 + \frac{1}{b^2}\left(-\frac{1}{2}(x^2 + (a + b)^2) + (a + b)x\right);\ a \le x \le a + b$$

Moments: $\mu = a, \sigma^2 = \frac{b^2}{6}, \gamma_1 = 0, \gamma_2 = 2.4$

Parameter Estimation: $\hat{a} = \bar{X}, \hat{b} = \sqrt{6}S = \sqrt{6}\left(\frac{1}{m-1}\sum_{i=1}^{m}(x_i - \bar{x})^2\right)^{1/2}$

***Uniform Distribution*:**

PDF: $f(x) = \frac{1}{b-a}, a \le x \le b$

Where a and b are the paremters of the distribution.

CDF: $F(x) = \frac{x-a}{b-a}$

Moments: $\mu = \frac{a+b}{2}, \sigma^2 = \frac{(b-a)^2}{12}, \gamma_1 = 0, \gamma_2 = 1.8$

Parameter Estimation: $\hat{a} = \overline{X} - \sqrt{3}S \; \hat{b} = \overline{X} + \sqrt{3}S = \overline{X} + \sqrt{3}\left(\frac{1}{m-1}\sum_{i=1}^{m}(x_i - \bar{x})^2\right)^{1/2}$

Sampling Methods

Sampling is the process that randomly generates possible values from input probability distribution functions. Each set of samples represents a possible combination of input values. With sufficient iterations of sampling, the samples form a distribution that approximates the known input probability distribution function. The number of iterations that is required to accurately replicate an input distribution through sampling is a major factor that evaluates different sampling techniques. Monte Carlo sampling (MCS) and Latin Hypercube sampling (LHS) are widely used sampling methods (Swiler and Wyss 2004).

Monte Carlo Sampling

MCS is the widely-used traditional technique. MCS uses random or pseudo-random numbers to sample from a probability distribution. In each sampling, a random number between 0 and 1 is generated and applied to select a value from a cumulative distribution function that is based on the input probability distribution. Sufficient sampling is performed to ensure replication of the input probability distributions. Clustering, however, could happen if a small number of iterations are carried out, especially when a distribution has low-probability outcomes.

Latin Hypercube Sampling

LHS is a stratified sampling technique. Stratification divides the distribution of each input variable into *n* non-overlapping intervals with equal probability. A sample is then randomly selected from each interval. The number of intervals is equal to the number of iterations. Once an interval is sampled, it is not sampled again. There are *n* values selected for each variable (X). To maintain independence between variables, it is necessary to randomly select the interval to make a sample for each variable. The *n* values of X1 are paired randomly with the *n* values of X2. The *n* pairs are further combined randomly with the *n* values of X3 to form *n* triplets, and so on, until *n* k-dimension vectors are formed. Sampling finally yields an $(n \times k)$ input matrix. There are two variants of LHS: Random LHS and Median LHS. Random LHS selects a random point within each interval of the given input distribution, whereas Median LHS uses the median of each interval. In comparison with MCS, LHS has better sampling efficiency in replicating input probability distributions and makes low-probability outcomes better represented in input probability distributions.

Halton/Hammersley Sampling

Discrepancy is used to measure how the samples deviate from a regular distribution. While random sequences may have a high discrepancy, a sequence of uniformly distributed samples would have a low discrepancy. Halton/Hammersley sampling is one of the low-discrepancy sampling methods and is an extended version of the van der Corput sequence.

Stochastic Simulation in the IECM

Figure 2 conceptually illustrates the procedure of uncertainty analysis in the IECM:

1) Uncertain input parameters are first identified and then specified with appropriate probabilistic distribution functions.
2) With identified uncertain parameters, MCS or LHS is performed for their assigned probability distributions in an iterative fashion to drive the technical model and then yield a range of output results, which demonstrate possible collective effects of simultaneous variations in many uncertain input parameters.
3) For a given variable, a cumulative distribution function is generated for output results to provide the likelihood of any particular value.

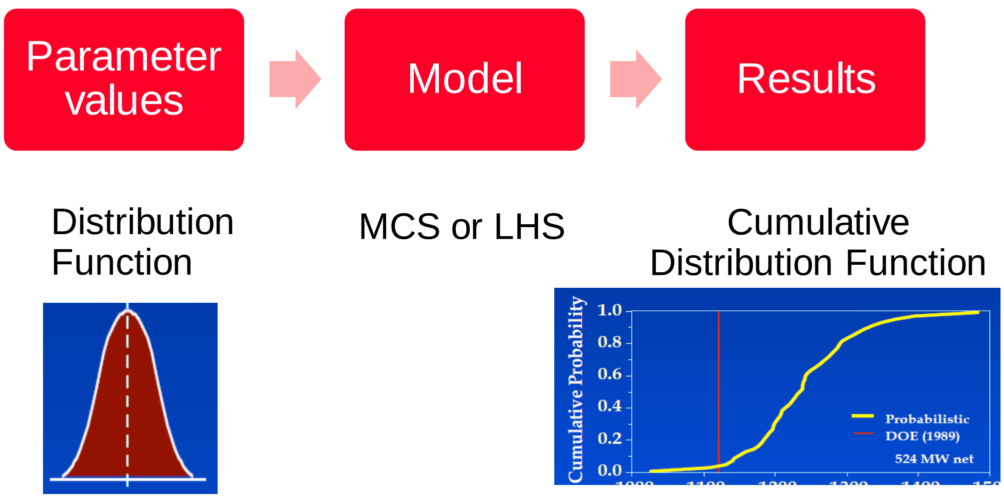
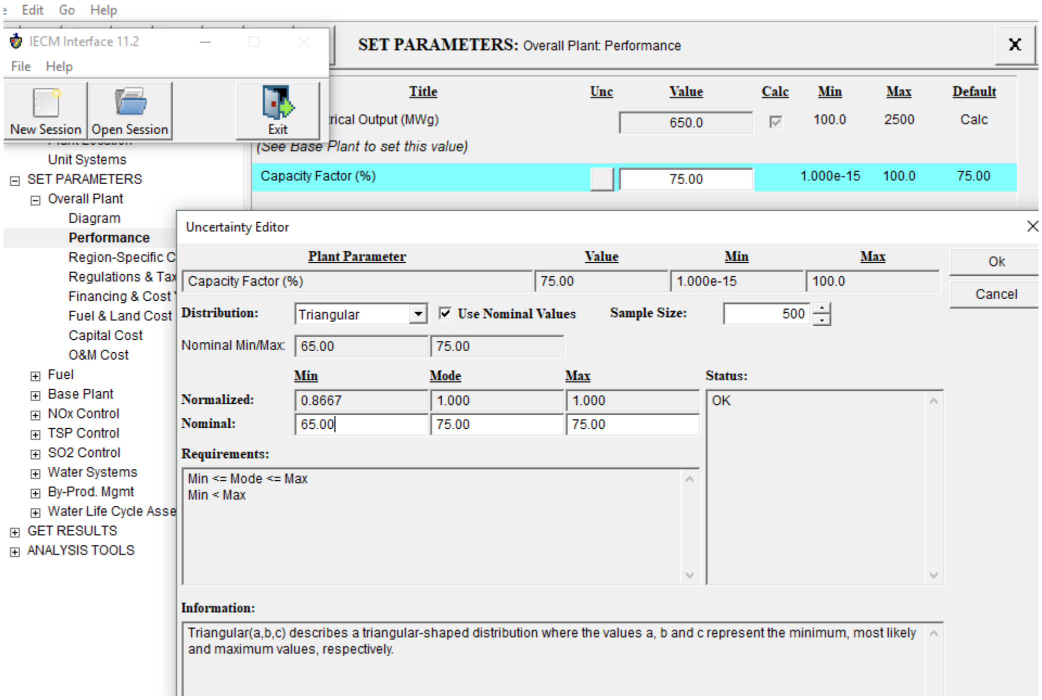


Figure 2. Procedure of Probabilistic Simulation

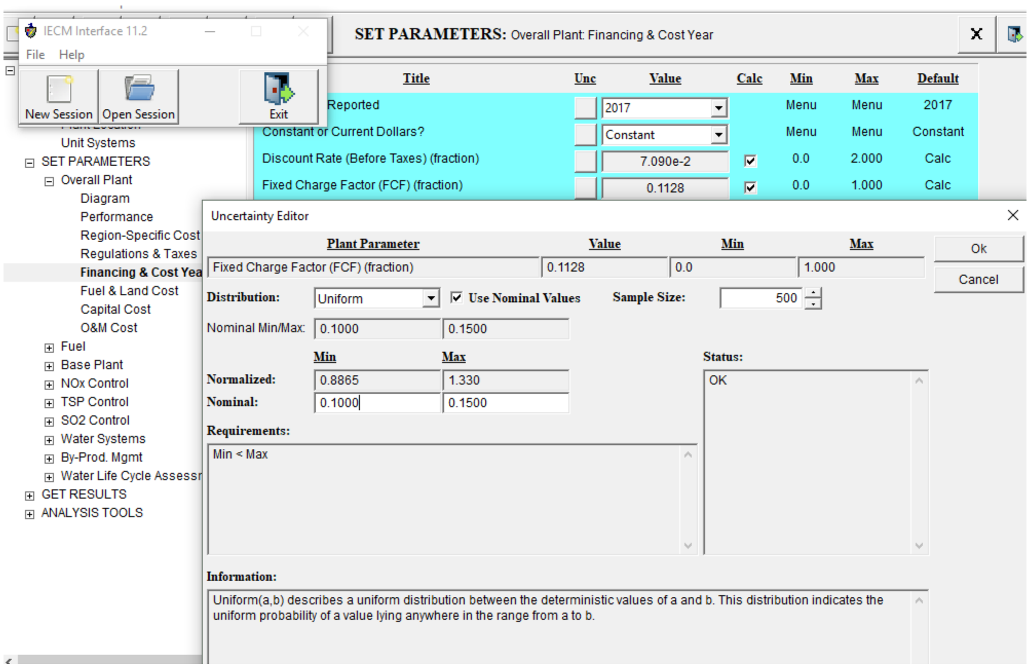
An Illustrative Case Study Using the IECM

Question: To determine the cumulative distribution of the cost of electricity generation of a coal-fired power plant in the IECM that meets new source performance standards, in which capacity factor and fixed charge factor are uncertain, and other parameter values are in default.

To address it, there are three major steps involved as follows:
Step 1: Identify uncertain parameters and then assign probability distribution functions:



(a) Uncertain Capacity Factor



(b) Uncertain Fixed Charge Factor

Figure 3. Assign Probability Distribution Functions to Uncertain Capacity Factor and Fixed Charge Factor in the IECM

Step 2: Select sampling method and sample size and then perform stochastic simulation:

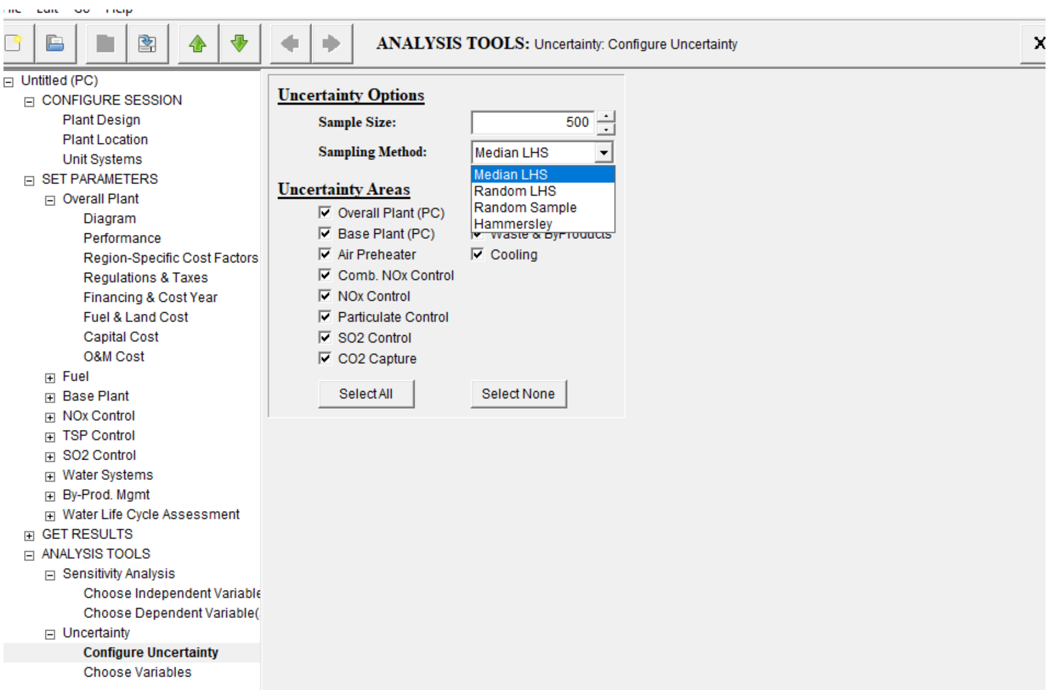


Figure 4. Select Sampling Method and Sample Size in the IECM

Step 3: Yield a cumulative distribution function and estimate likelihood of specific outcomes:

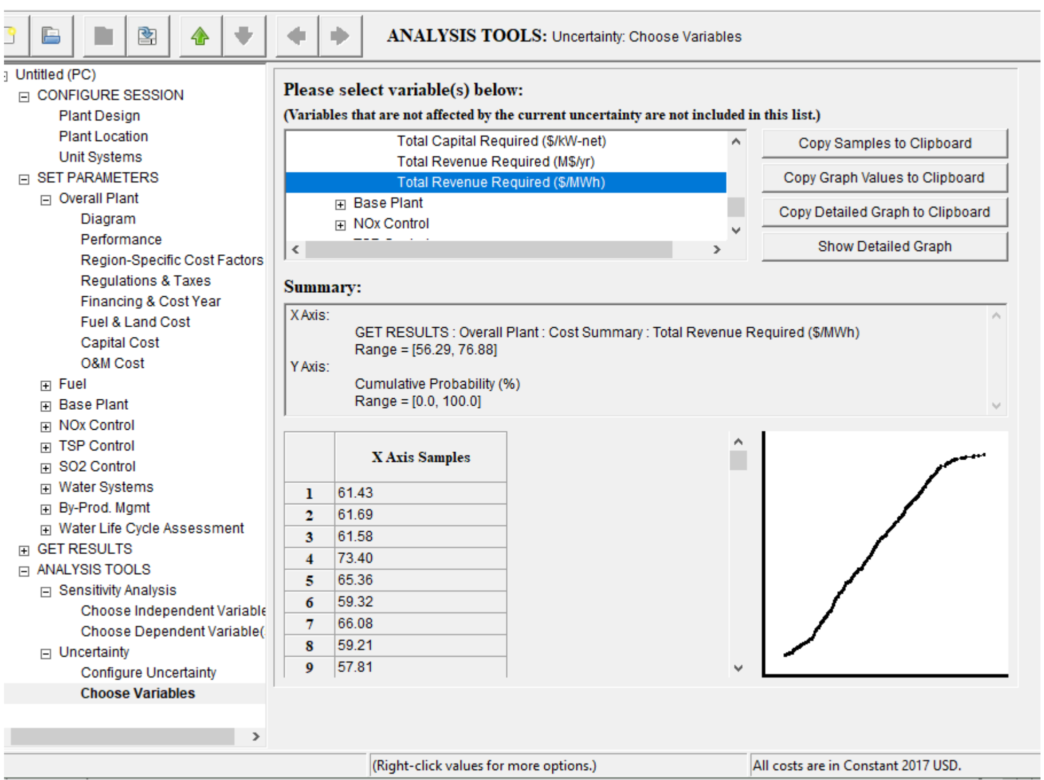


Figure 5. Yield a Cumulative Distribution Function in the IECM

References

Diwekar, U. M., & Rubin, E. S. (1991). Stochastic modeling of chemical processes. *Computers & Chemical Engineering*, *15*(2), 105-114.

Morgan, M. G., Henrion, M., & Small, M. (1992). Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. Cambridge university press.

Swiler, L.P., & Wyss, G.D. (2004). A User's Guide to Sandia's Latin Hypercube Sampling Software: LHS UNIX Library/Standalone Version; No. SAND2004-2439, Sandia National Laboratories, Albuquerque, NM.