STOCHASTIC MODELING OF CHEMICAL PROCESSES

U. M. Diwekar, H. C. Frey, and E. S. Rubin
Carnegie Mellon University
Pittsburgh, PA 15213

INTRODUCTION

The increasing capability of computer processing has facilitated the development and use of software to simulate chemical and other industrial processes. Steady state processes are well served in this area by process simulators such as FLOWTRAN, PROCESS, CHEMSHARE, and ASPEN. Several companies have in-house simulation packages catering to their plant needs. However, while computer simulation models have grown in sophistication, certain capabilities important to research and development planning, system design, and economic analysis still remain to be developed.

The conventional simulation models or simulators typically employ a Fortran code which produces deterministic (point estimate) results for a particular set of input assumptions. Such an approach can be overly simplistic, however, often resulting in the addition of large safety or "fudge" factors to accommodate the reality of uncertainty in equipment or process design. Chemical plants are usually faced with uncertain conditions during their operation. These uncertainties can arise from variations either in external parameters, such as the quality of feed streams, or from internal process parameters such as transfer coefficients, reaction constants, and physical properties. If the technology is new, there are additional uncertainties due to limited performance data.

The ability to analyze uncertainty is especially important in the context of ongoing research and development, where technical and economic parameters for individual processes and system designs are not well established. Uncertainties also are important in comparing advanced system designs with "baseline" systems reflecting currently commercial technology. To analyze uncertainty, the capability to perform sensitivity analysis through a series of multiple runs is usually available. Typically, however, only one or two parameters at a time are varied in a simulation framework which may contain a large number of independent variables. Thus, important interactions or cases may be overlooked. Although larger number of cases may be run as part of a sensitivity study, the volume of output that is generated makes results cumbersome or difficult to interpret and/or display. Even where many cases are analyzed, sensitivity analysis still provides no information as to the likelihood of different outcomes. In short, the process

analysis of real systems requires both stochastic and deterministic modeling capabilities.

Freeman and Gaddy (1975) and some earlier workers [e.g., Lashmet and Szczepanski (1974), Kittrell and Watson (1966)] applied probabilistic techniques for the calculation of overdesign factors. Error analysis and modeling of a reactor using statistical techniques is reported by Atherton et al.(1975). U.S. federal agencies have also committed to the continued development and use of probabilistic studies for risk analysis. One of the earliest large-scale studies was the nuclear reactor safety study generally known as the "Rasmussen Report" [Rasmussen et al.(1975)].

The problem of flexibility of design and the concept of a "flexibility index" have been discussed in recent years [Grossmann et al.(1983), Morari (1983), Pistikopoulos and Grossmann (1988)] . Application of optimization techniques to obtain flexibility of operation under uncertainty is the objective of most of these studies. Use of discrete simulation languages like GPSS for batch processing scheduling is another problem which has received considerable attention in the current literature [Schultheisz and Sommerfeld (1988)]. Batch processing involves both state and time events. The stochastic modeling capability based on queuing theory is used to solve the problem of scheduling of time events in batch processing.

Though uncertainty analysis techniques are gaining attention in the literature, none of the well-known process simulators have the capability to handle uncertainties systematically. This paper discusses a new generalized stochastic modeling capability built around the ASPEN process simulator and illustrates its use for a complex process.

THE STOCHASTIC SIMULATION MODEL

The approach adopted involves adding a stochastic modeling capability for uncertainty analysis to the public version of the ASPEN simulator developed for the U.S. Department of Energy (DOE) (MIT 1982). To implement the stochastic modeling capability, ASPEN's modular nature (consisting of unit operation modules or blocks) has been utilized. The stochastic simulation module is based on the public domain programs of Iman et al. [Iman and Shortencarier (1984), Iman et al.(1985)]. A new unit operation block, called STOCHA, has been added to the ASPEN unit operation module library. Initially, this new block was added in the form of a user block. More recently, it has been added as a permanent unit operation block in ASPEN. The structure of the block and it's use are briefly described below. Details are provided elsewhere [Diwekar (1989)].

The unit operation block, STOCHA, characterizes the uncertainty in model input parameters in terms of probability distributions, and analyzes their effect on selected output variables. To link STOCHA to the ASPEN flowsheet, two Fortran blocks are needed. This type of stochastic

modeling capability can be used for systematic probabilistic analysis. The stochastic modeling approach involves:

1. Specifying the uncertainties in key input parameters in terms of probability distributions;
2. Specifying the correlation structure of any inter-dependent parameters;
3. Sampling the distributions of the specified parameters in an iterative fashion;
4. Propagating the effects of uncertainties through the process flowsheet; and
5. Applying graphical and statistical techniques to analyze the results.

The software provides a set of eight types of probability distributions for characterizing input variables (normal, lognormal, uniform, loguniform, modified uniform, modified loguniform, beta, and triangular distributions). There is also the ability to specify any user-defined distribution. The stochastic block assigns user-specified distributions to the flowsheet input parameters, then uses either Latin Hypercube sampling (LHS) or random sampling to pass the sampled values of each uncertain variable to the flowsheet. After a flowsheet simulation is run, the output variables of interest are collected. The simulation is then repeated for a new set of samples selected from the probabilistic input distributions. A new Fortran block is used to control the cycling of the stochastic block, and another Fortran block, called STCTAIL, is used to access and assign samples to model parameters.

After all samples or observations have gone through this cycle for a specified number of times (typically 20 to 100 or more, depending on the accuracy sought by the user), the stochastic block analyzes the output. The output options include cumulative and non-cumulative probability density functions of input and output parameters, plus multivariate correlation coefficients which measure the effects of input uncertainties on the output results.

Figure 1 shows the use of the stochastic block for uncertainty analysis of a flowsheet. The cycle for uncertain variables consists of: (a) the stochastic block, STOCHA, for assigning parameter uncertainty distributions; (b) the Fortran block, STCTAIL, for accessing variables and assigning sampled values; and (c) the Fortran recycle block, STCREC, for data output collection and recycling.

If there are convergence loops nested within the flowsheet of a stochastic cycle, and if entire flowsheet calculations are involved in the analysis, it may be necessary to use an additional Fortran cycle control block to mark the beginning and end of the convergence cycle. Appropriate software also has been implemented to handle this case, as described by Diwekar (1989).

STATISTICAL TERMS AND HEURISTICS

The uncertainties or variability in engineering models can be expressed in terms of probabilistic distributions. The probability distributions show the range of values the variable could take and the likelihood of occurrence of each value within the range. Thus, the distributions define the rule for describing the probability measures associated with the values of a random (uncertain) variable. Probability distributions may be described in their entirety as cumulative distribution functions, or by selected parameters, such as fractiles or moments (e.g. mean and variance). More complete review of these methods may be found in Diwekar and Rubin (1989).

Specifying a Probability Distribution

To accommodate the diverse nature of uncertainty, the different distributions available within the stochastic block and their features are shown in Figure 2. The type of distribution chosen for an uncertain variable reflects the amount of information available. The uniform and loguniform distributions represent an equal likelihood of a value lying anywhere within a specified range, on either a linear or logarithmic scale, respectively. The modified forms of these distributions, uniform* and loguniform*, allow several intervals of the range to be distinguished.

A normal (Gaussian) distribution reflects a symmetric but varying probability of a parameter value being above or below the mean value. In contrast, lognormal and triangular distributions are skewed such that there is a higher probability of values lying on one side of median than the other. A beta distribution provides a wide range of shapes and is a very flexible means of representing variability over a fixed range. Finally, user-specified distributions can be used to represent any arbitrary characterization of uncertainty, including chance constraints (i.e., fixed probabilities of discrete values).

Sampling

There are two sampling methods available with the stochastic unit operation block: random Monte Carlo sampling and Latin Hypercube sampling (LHS). In the LHS method, a distribution is divided into intervals of equal probability, and one sample is taken at random from within each interval. LHS guarantees that the values from the entire range of the distribution are sampled in proportion to the probability density of the distribution, whereas in traditional Monte Carlo methods samples are taken at random. In LHS, the order of sampling is random, but the entire distribution is sampled. Because the distributions are sampled over the entire range of probable values, the number of samples required to adequately represent a distribution is less than for Monte Carlo simulations. The procedure to select samples using Latin Hypercube sampling is described in Diwekar and Rubin (1989). Since LHS is generally more efficient than Monte Carlo, the primary reason for providing the latter method is that the statistics may be harder to compute with LHS. With standard LHS, the sample scenarios and outputs are random but they are not completely independent. Thus, statistics for estimating the precision of the results (e.g.,

confidence intervals) may be underestimated, though the level of precision remains higher.

The sampling procedure (LHS or random sampling) is applied to each input variable. The final step in sampling involves pairing the selected values for correlated variables. This pairing is done either randomly or using the restricted technique of Iman and Conover (1982), where user-specified correlations are used. If a correlation structure is not specified by the user, then the model computes a measure for detecting large pairwise correlations. This measure is known as the variance inflation factor (VIF), defined as the largest element on the diagonal of the inverse of the correlation matrix. The VIF appears in the model output report when the user requests the input correlation matrix to be printed.

Choosing Sample Size

The unit operation block STOCHA provides a measure of the precision of an estimate of the cumulative distribution function in terms of a confidence interval. The α confidence interval for Y_p, the p-th fractile, is given by (y_i, y_k), where y_i is the lower end point of the interval, y_k is the upper end point, and

$$i = np - c\sqrt{np(1-p)} , \qquad k = np + c\sqrt{np(1-p)} \qquad (1)$$

where n is the sample size and c is the deviation enclosing probability α of the unit normal distribution (obtained by inverting the normal distribution function, which is approximated by the multiplication of two polynomials). [Griffiths and Hill (1985)]. From the values of the confidence interval precision, the sufficiency of the number of samples, n, can be judged.

Sensitivity Analysis

Once the input samples have passed through the flowsheet and all the sample runs are completed, the stochastic block can be used to quantify the sensitivity of an output to each input parameter. Two closely related but different measures are presented. These are partial correlation coefficients and standardized regression coefficients.

From the sampling data it is possible to construct an approximate regression model which relates an output parameter y_i to the input parameters x_i:

$$y_i = b_0 + \sum_j b_j x_j \qquad (2)$$

The constants b_j are ordinary regression coefficients which are easily influenced by units of measurement. This problem can be circumvented if the regression model can be written using the transformed variables x* and y* given by:

$$x^* = \frac{(x - \mu_x)}{\sigma_x}, \qquad y^* = \frac{(y - \mu_y)}{\sigma_y} \tag{3}$$

and a regression model in the standardized form:

$$y^* = \sum_j b_j^* x_j^* \tag{4}$$

The coefficients in this model are called standardized regression coefficients and they provide a direct measure of the relative importance of the input variables. The accuracy of this model can be judged by the value of R_y^2, the coefficient of determination, which is:

$$R_y^2 = \frac{\sum_i (\hat{y}_i - \mu_y)^2}{\sum_i (y_i - \mu_y)^2} \tag{5}$$

where $\hat{y}$ is the calculated value of y_i using the regression model.

The partial correlation coefficients provides a measure of the linear relationship between the output and input variables. If the correlation is denoted by $\left| r_{yj} \right|$, then the maximum absolute value can be used to identify the input variables having the strongest input relationship. When nonlinear relationships are involved, the standardized correlation coefficients and partial regression coefficients are calculated based on ranks rather than absolute value.

ILLUSTRATIVE RESULTS

To illustrate the type of results obtainable with the new ASPEN stochastic block, one specific Integrated Coal Gasification Combined Cycle Power Plant flowsheet is considered.

A Complex Flowsheet: Integrated Gasification Combined Cycle Power Plant

The Illustrative example analyzes an integrated gasification combined cycle (IGCC) power system, which is a promising new approach for the clean and efficient use of coal for electric power generation. At the present time, however, there is still little experience with IGCC systems on a commercial scale. The uncertain nature of the limited performance and cost data for first generation systems, coupled with uncertainties associated with alternative process configurations, suggests a strong need for a systematic analysis of uncertainties in IGCC process

design.

ASPEN performance models developed for IGCC systems by the U.S. Department of Energy (DOE) include different gasifier designs and gas stream cleanup systems, including models of fixed bed, fluidized bed and entrained bed gasifiers, plus cold gas and hot gas cleanup systems [Stone (1985)]. These ASPEN models typically consist of approximately 80 unit operation blocks and eight flowsheet sections. While the bulk of the models are comprised of generalized unit operation blocks (e.g., pumps, heat exchangers, pressure vessels, etc.), there are a number of Fortran blocks and design specification blocks which are specific to IGCC systems, or to a particular flowsheet. There are also user-specified models to handle coal properties, and a Fortran block used as a summary report writer.

Use of the probabilistic modeling capability is illustrated using a simulation model based on a conceptual design for a commercial IGCC plant using the Kellogg-Rust-Westinghouse (KRW) ash agglomerating pressurized fluid-bed gasifier with conventional cold gas cleanup [Bechtel (1983)]. ASPEN performs a steady-state computer simulation of this process. The simulation flowsheet contains 83 unit operation blocks (i.e., models of unit operations such as pumps, reactors, and compressors), plus user routines to calculate ash enthalpy and coal decomposition. The eight plant sections which are simulated are: gasification, solids separation, gas cleaning and cooling, gas turbine, effluent water primary treatment system, Claus plant, Beavon-Stretford unit, and the steam cycle.

The KRW-IGCC simulation model was developed by DOE to provide accurate mass and energy balances, and to track major environmental species (SOx, NOx, and particulates). The overall simulation runs in 200 seconds of CPU time on a Vax-3200 computer workstation. Table 1 shows a summary of results obtained for a case study of a 678 MW system using deterministic input parameters [Stone (1985), Fluor (1985)].

For this example, six input variables in the KRW flowsheet are treated as uncertain. Table 2 shows the deterministic values of these parameters as well as the assumed uncertainties, given as probability distributions. The distributions assigned to the variables in Table 2 are based on data available from the current literature, and reflect the variability in judgments of different design teams in selecting parameter values for similar IGCC systems. For this example, most input parameters have been assigned a skewed triangular distribution while the uncertainty of one parameter is represented by a uniform distribution. All parameters are taken to be independent, as supported by the technical literature.

The effect of these uncertainties on selected model output parameters is shown in Figures 3 and 4 in the form of cumulative probability distributions (CDFs) based on 200 samples. The information obtained from these distributions and associated statistical data (not shown here)

provide insights not available from conventional deterministic analysis (e.g., the range of uncertainty of an output variable of interest; the most likely value; the range encompassed by a given confidence level; the chances that the value will be below or above some certain level).

For the KRW IGCC flowsheet, Figure 4 shows the effect of input uncertainties on the overall plant efficiency. This result shows only a 3% chance of achieving the expected (deterministic) performance, with a 5% chance that the efficiency will be below 29.22% or above 36% (which defines the 90 percent confidence level). Combinations of independent input parameter values leading to high thermal efficiency can be found from a more detailed examination of the probabilistic results. Thus, a probabilistic analysis also can help identify globally optimum conditions for IGCC system designs.

Apart from uncertainty analysis results, the stochastic modeling framework also displays the effect of different input variables on output results in the form of correlation coefficients. For example, the sensitivity of the overall plant efficiency based on the correlation coefficients found for this case study (Table 3) are, in descending order of importance: (1) fraction of Carbon going to gasifier, (2) gas turbine pressure ratio, (3) gas turbine firing temperature, (4) recycle split, (5) fraction of Sulfur going to gasifier, and (6) fraction of H2S in Acidgas.

CONCLUSION

This paper has described a new stochastic modeling capability for the ASPEN chemical process simulator. The stochastic modeling capability can be used to evaluate the performance of any chemical plant which can be formulated using the simulator. Applications of this capability to a complex flowsheet also was illustrated. The generalized stochastic modeling capability is useful for performance analysis, economic analysis, comparisons of different technologies, determination of overdesign factors, error and sensitivity analysis, risk analysis, feasibility studies, identification of process R&D priorities, and research management and planning. The methodology is universal and can be easily extended to other simulators or modeling packages.

ACKNOWLEDGEMENTS

REFERENCES

Atherton R.W., R.B. Schainker, and E.R. Ducot, *AIChE J.*, 21, 144 (1975).

Bechtel Group, Inc. 1983. Design of Advanced Fossil-Fuel Systems (DAFFS): A Study of Three Developing Technologies for Coal-Fired, Base-Load Electric Power Generation. Integrated Coal-Gasification Combined-Cycle Power Plant with Westinghouse Gasification Process. ANL/FE-83-17. NTIS/DE83017731.

Diwekar, U.M. 1989. User's Manual For Stochastic Simulation in ASPEN. Prepared by Center for Energy and Environmental Studies, Carnegie Mellon University, Pittsburgh, PA, for U.S. Department of Energy, Morgantown Energy Research Center, Morgantown, WV, May 1989.

Diwekar, U.M and E.S. Rubin. 1989. Technical Reference Manual For Stochastic Simulation in ASPEN. Prepared by Center for Energy and Environmental Studies, Carnegie Mellon University, Pittsburgh, PA for U.S. Department of Energy, Morgantown Energy Research Center, Morgantown, WV, May 1989.

Fluor Engineers, Inc. 1985. Cost and Performance of Kellog Rust Westinghouse-based-Gasification-Combined-Cycle Plants. Final report, EPRI-AP-4018

Freeman R.A. and J.L.Gaddy, *AIChE J.*, 21, 436 (1975).

Griffiths P, and I. D. Hill 1985. *Applied Statistics Algorithms.* Ellis Horwood Ltd. Chichester, West Sussex, England, first ed 1985.

Grossman I.E., K.P. Helemane, and R.E. Swaney, *Comput. Chem.Engng.*, 7, 43 (1983).

Iman R.L. and Conover W.J., *Commun. Statist.-Simula. Computa.*, 11(3), 331 (1982).

Iman, R.L., and M.J. Shortencarier 1984. A Fortran 77 Program and User's Guide for the Generation of Latin Hypercube and Random Samples for Use with Computer Models. Sandia National Laboratories, Albuquerque, New Mexico. SAND83-2365, March 1984.

Iman, R.L., M.J. Shortencarier, and J.D. Johnson 1985. A Fortran 77 Program and User's Guide for the Calculation of Partial Correlation Coefficients and Standardized Regression Coefficients. Sandia National Laboratories, Albuquerque, New Mexico. SAND85-0044, June 1985.

Kittrell J.R. and C.C. Watson, *Chem. Engg. Prog.*, 62, 79 (1966)

Lashmet P.K. and S.Z. Szczepanski, *Ind.Eng.Chem.Pro.Des.Dev.*, 13, 103 (1974).

Massachusetts Institute of Technology 1982. ASPEN User's Manual, Volume 1, DOE/MC/16481-1203. NTIS/DE82020196.

Morari M., *Comput. Chem.Engng.*, 7, 423 (1983).

Pistkipoulos E.N. and Grossman I.E., *Comput. Chem.Engng.*, 12, 841 (1988).

Rasmussen N. 1975, Reactor Safety Study, Technical Report WASH-1400, US Nuclear Regulatory Commission 1975.

Schultheisz D.J. and J.T. Sommerfeld 1988, Exercises in Chemical Engineering Using GPSS, CACHE CORPORATION, austin, Texas, December 1988.

Stone, K.R. 1985. ASPEN Simulations of Fluidized-Bed and Entrained-Flow Integrated Gasification Combined-Cycle Power Plants. Technical Note, DOE Morgantown Energy Research Center, Morgantown, WV, July 1985.

TABLE 1. KRW FLUIDIZED-BED IGCC POWER PLANT SYSTEM SUMMARY

*** GASIFIER CONDITIONS ***

Coal Flow Rate: 0.567703E+06 lbs/hr

Oxygen Flow Rate: 0.386038E+06 lbs/hr

Steam Flow Rate: 0.346299E+06 lbs/hr

Gasifier Pressure: 465.0 psia

Gasifier Temperature: 1850 F

*** MS7000 GAS TURBINE CONDITIONS ***

Fuel Flow Rate: 0.874669E+06 lbs/hr

Air Flow Rate: 0.899985E+07 lbs/hr

Steam Flow Rate: 0.262401E+06 lbs/hr

Fuel LHV: 5185.4 Btu/lb, 272.9 Btu/scf

Fuel HHV: 5600.9 Btu/lb, 294.8 Btu/scf

Firing Temperature: 2300.0 F

Combustor Exit Temperature: 2425.0 F

Turbine Exhaust Temperature: 1160.4 F

Thermal Efficiency (LHV) : 0.3484

Generator Efficiency: 0.9850

*** STEAM TURBINE CONDITIONS ***

Superheated Steam Flow Rate: 0.164041E+07 lbs/hr

Superheated Steam Temperature: 984.7 F

Expanded Steam Quality: 0.9516

Generator Efficiency: 0.9850

*** POWER PRODUCTION SUMMARY ***

Gas Turbine: 0.463128E+09 Watts

Steam Turbine: 0.290632E+09 Watts

Compressors: -0.731738E+06 Watts

Pumps: -0.359428E+07 Watts

Oxygen Plant: -0.717608E+08 Watts

Plant Total: 0.677673E+09 Watts

****

PLANT THERMAL EFFICIENCY (HHV) = 0.3623

****

TABLE 2. ASSUMED UNCERTAINTY IN MODEL INPUT PARAMETERS

Parameter (Units)	Nominal Value	Distribution	Range (Mode)
Fraction of Carbon going to gasifier	0.95	Uniform	0.75 - 0.95
Fraction of Sulfur going to gasifier	0.99	Triangular	0.8 - 0.99 (0.8)
Recycle split	0.17	Uniform	0.12 - 0.17
Gas turbine pressure ratio	12	Triangular	12 - 13.5 (13.5)
Gas turbine firing temp.	2300	Uniform	2200 - 2300
Fraction of H2S in Acidgas	0.935	Uniform	0.935-0.99

TABLE 3. SENSITIVITY ANALYSIS RESULTS IN TERMS OF PARTIAL CORRELATION FOR OVERALL PLANT EFFICIENCY

| Input Parameter	Partial Correlation Coefficient
Fraction of Carbon going to gasifier	-1.0
Gas turbine pressure ratio	0.878
Gas turbine firing temperature	0.850
Recycle split	-0.648
Fraction of Sulfur going to gasifier	-0.463
Fraction of H2S in Acidgas	-0.098

Figure 1. Basic Operation of the Stochastic Simulation Method
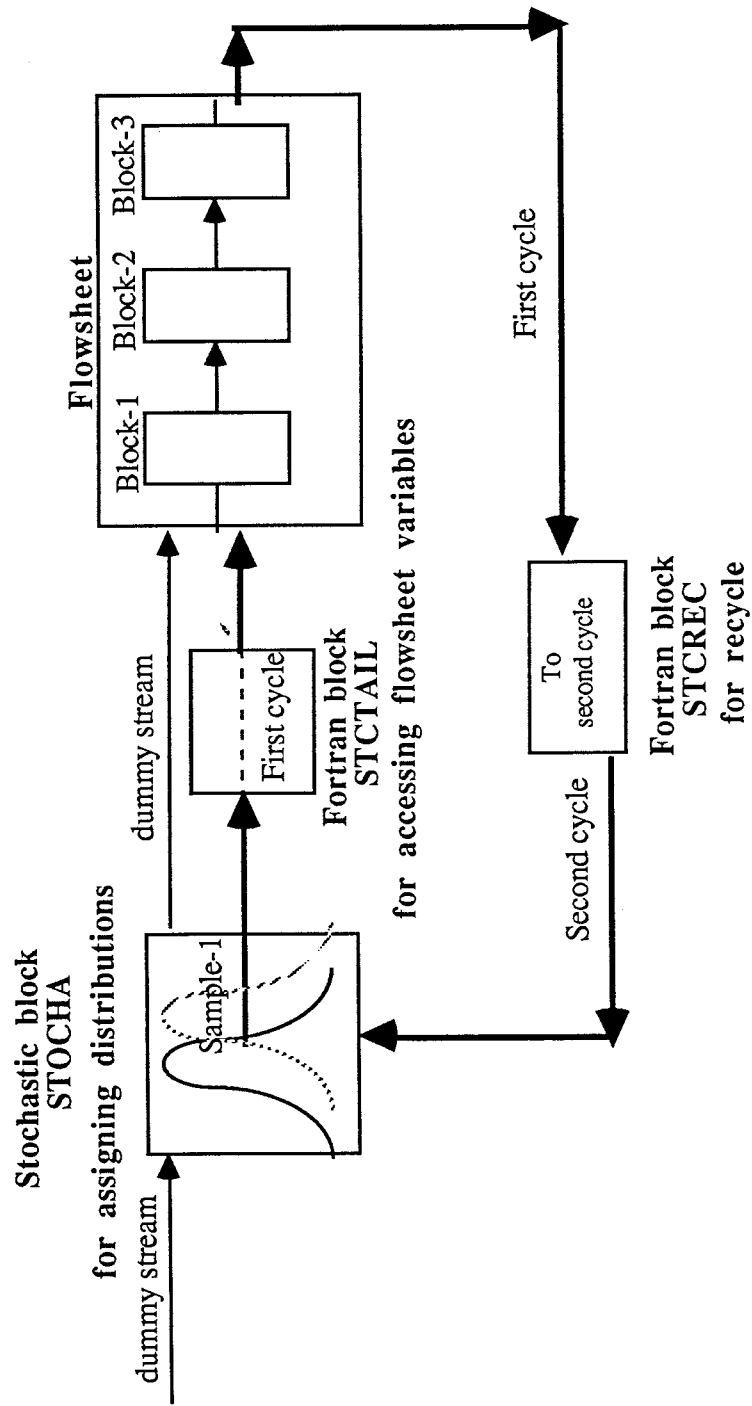
Figure 2. Available Distribution Functions for Model Input Parameters
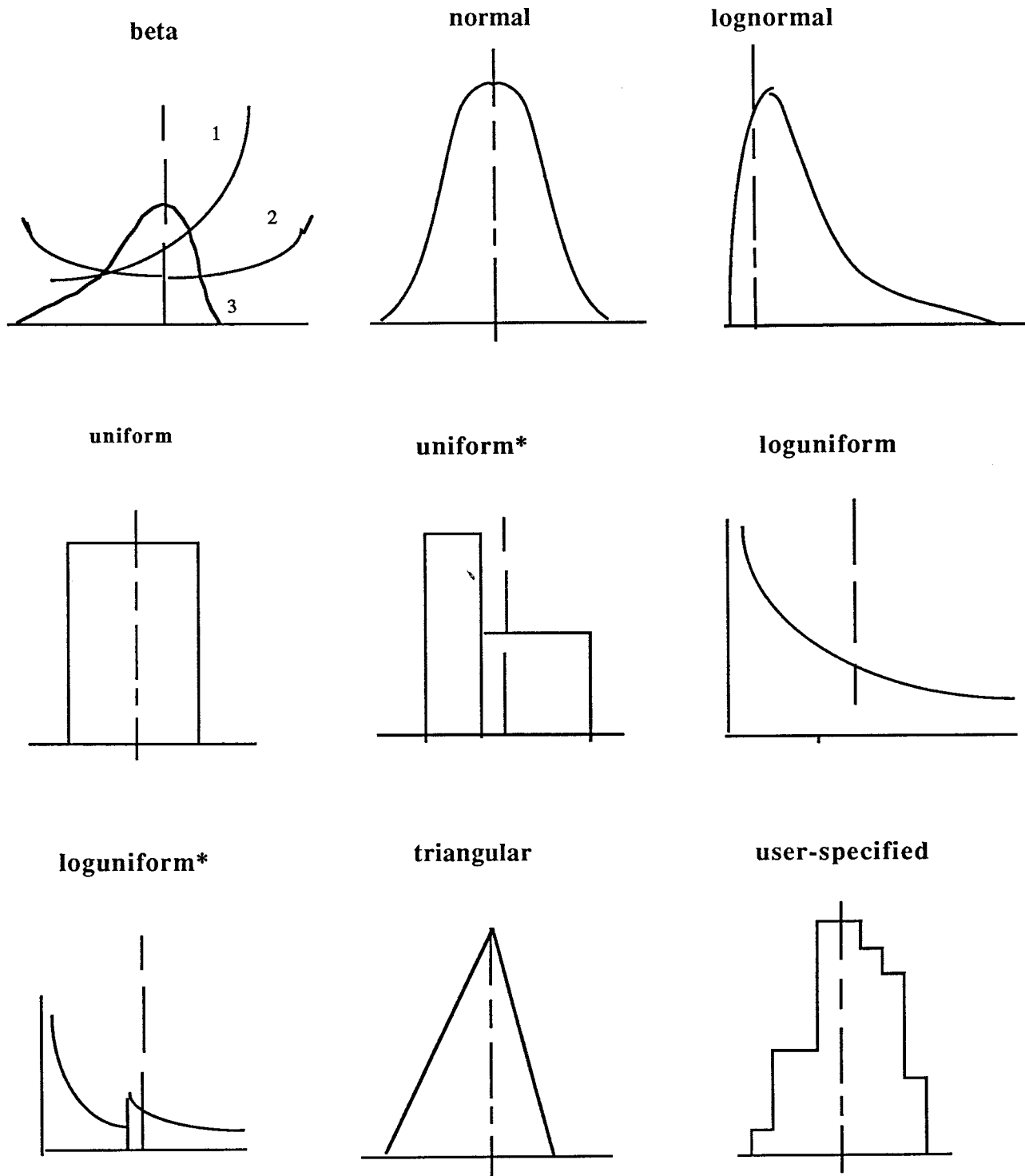
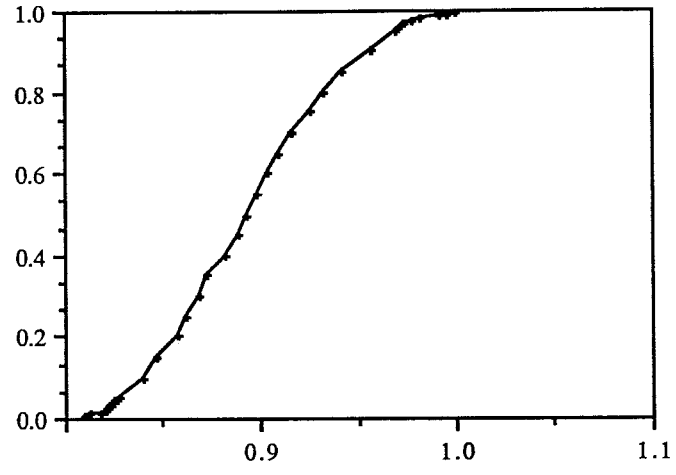# Figure 3. Cumulative Probability Distributions of Selected Output Parameters
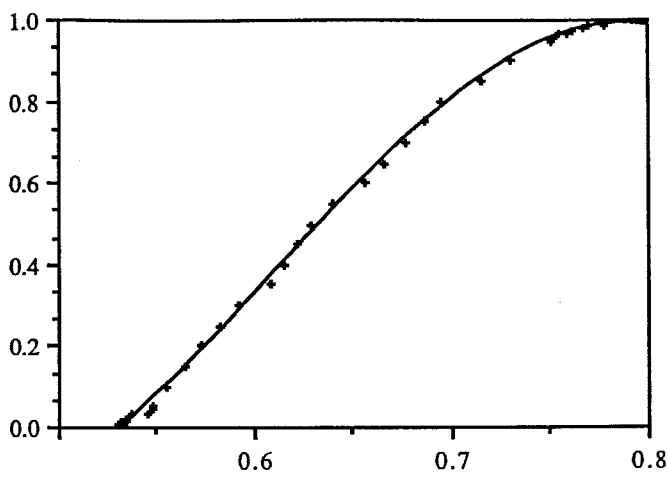
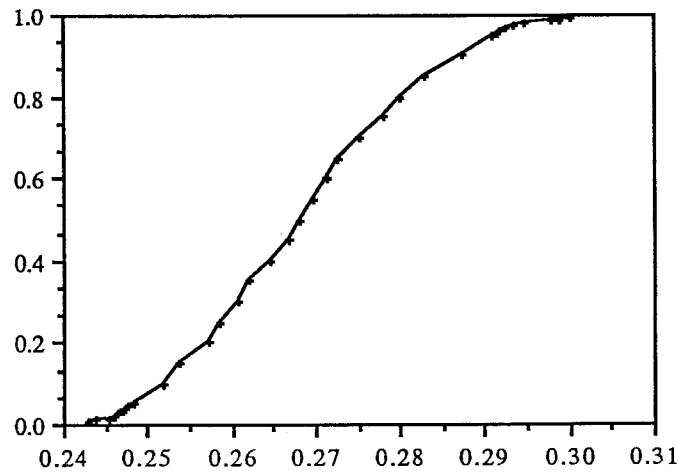**Coal Flowrate** (10^6 lbmoles/hr)
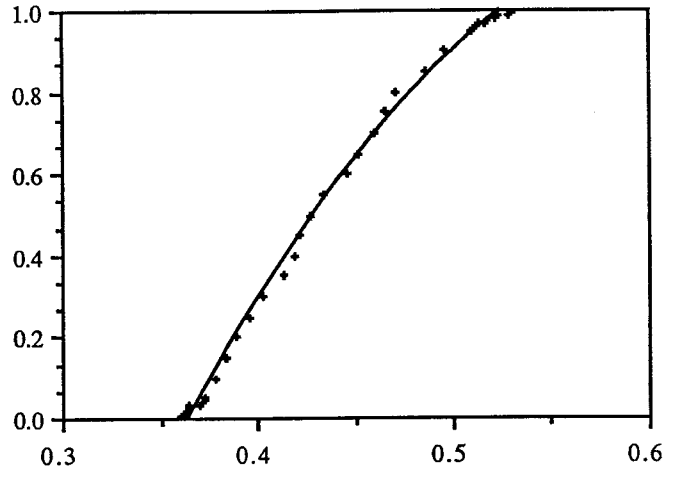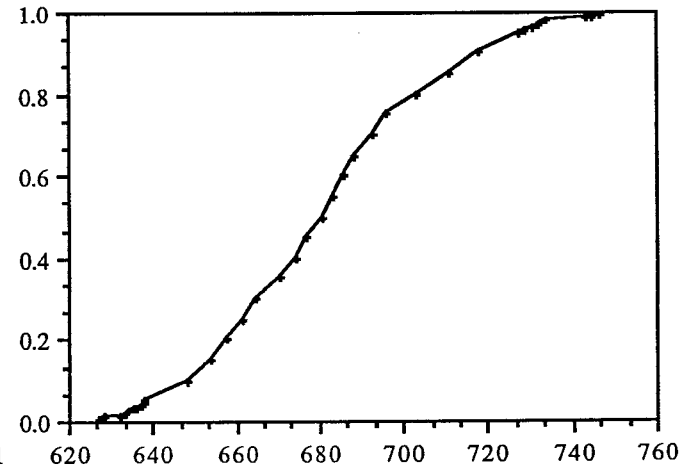
Oxygen Flowrate (10^6 lbmoles/hr)
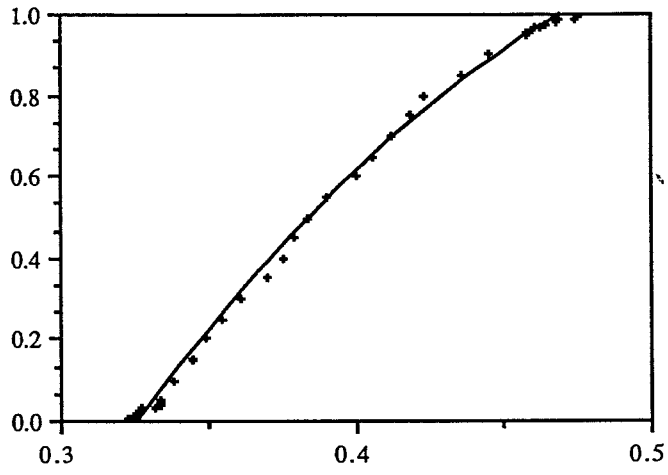
Gasifier Steam Flowrate (10^6 lbmoles/hr)

Fuel Flowrate (10^6 lbmoles/hr)

Gas Turbine Steam Flowrate (10^6 lbmoles/hr)

Total Power (MW)

Figure 4. Cumulative Probability Distribution of Total Plant Efficiency