

# ROLE OF SAMPLING IN PROCESS DESIGN, OPTIMIZATION AND CONTROL

*Saadet Ulas Acikgoz, Urmila M. Diwekar, University of Illinois at Chicago, Department of Bioengineering, Chicago, IL, Vishwamitra Research Institute, Westmont, IL*

## Abstract

Sampling is a statistical procedure used for uncertainty analysis and stochastic modeling in many applications related to process design, operation, control and optimization. The increased efforts for pollution prevention and sustainability necessitate a more contemporary approach for process design, in which objectives such as reducing environmental and health impacts, increasing reliability, safety and controllability, reducing risk and achieving better profitability are considered simultaneously. In this multifaceted approach to process design and operation, uncertainties need to be considered and included in process models and simulations. These uncertainties can be static or dynamic. A review of sampling techniques such as Monte Carlo sampling, importance sampling, Latin Hypercube Sampling (LHS) and Hammersley Sequence Sampling (HSS) are provided here. These sampling techniques have a wide range of use in chemical and process synthesis, process scheduling, supply chain management, process control and reliability as well as risk management. Apart from uncertainty analysis, sampling and sampling accuracy plays an important role in discrete, stochastic, and multi-objective optimization algorithms. This article also describes the recent developments in this area. In addition, future trends in process design and sampling techniques are also presented.

## Introduction

Sampling is a statistical procedure which involves the selection of a finite number of individuals to represent and infer some knowledge about a population of concern. Sampling techniques are used in wide range of science and engineering applications; they are of basic importance in computational statistics, in the implementation of probabilistic algorithms and in related problems of statistical computing that have a stochastic ingredient (e.g., financial modeling, artificial intelligence, computational chemistry, risk and uncertainty analysis and design of experiments). This article is devoted to role of sampling in process systems engineering.

Uncertainty analysis is a crucial step in process design and development due to the fact that over the life cycle of the process, product demands change, there may be variations in feedstock and product specifications and the process may be subject to short-term and long-term uncertainties. Furthermore, increased environmental consciousness in recent years and the efforts for pollution prevention necessitate chemical manufacturing plants to comply with stricter environmental regulations and to reduce waste. Therefore, the breadth of traditional process design approaches should be extended to include green engineering principles early in design (1, 2). This is because the decisions made earlier during the development of a chemical process affect later stages such as material and equipment selection, pilot plant studies and financial analysis and the opportunities for reducing environmental and health impacts of a process diminishes. Therefore, unlike traditional process design, where engineers are seeking only low cost options, contemporary process design approaches include environmental and health impacts, process performance indices such as risk, reliability, safety and flexibility as well as controllability and profitability into decision making. Sampling plays an important role in defining and

quantifying these objectives. Further, nowadays process design is just not restricted to process simulation but includes steps such as discovery, chemical synthesis on one end, and management, planning, and control on the other end. As the breadth of this design framework is extended, uncertainties in the model increase and efficient algorithms and tools are needed to address this problem. Sampling is an important component of these algorithms and tools.

Fig. 1 shows an overview of this integrated framework proposed (2), which applies green engineering principles at every stage of process design and development. The first stage in process development is discovery, where chemicals and materials are selected and synthesized in a laboratory or using computational chemistry methods. These methods use Monte Carlo methods based on sampling of the molecular configurational space.

Computer-aided molecular design (CAMD) is a commonly used technique for chemical synthesis where the reverse use of group contribution methods is employed to select materials with desired physical, chemical, environmental, and biological properties. The next stage of chemical synthesis is process synthesis where a chemical process is developed by choosing various unit operations and their connections. A flowsheet of the proposed plant is generated and process simulators are used to compute mass and energy flows for the process to predict its behavior if it was constructed.

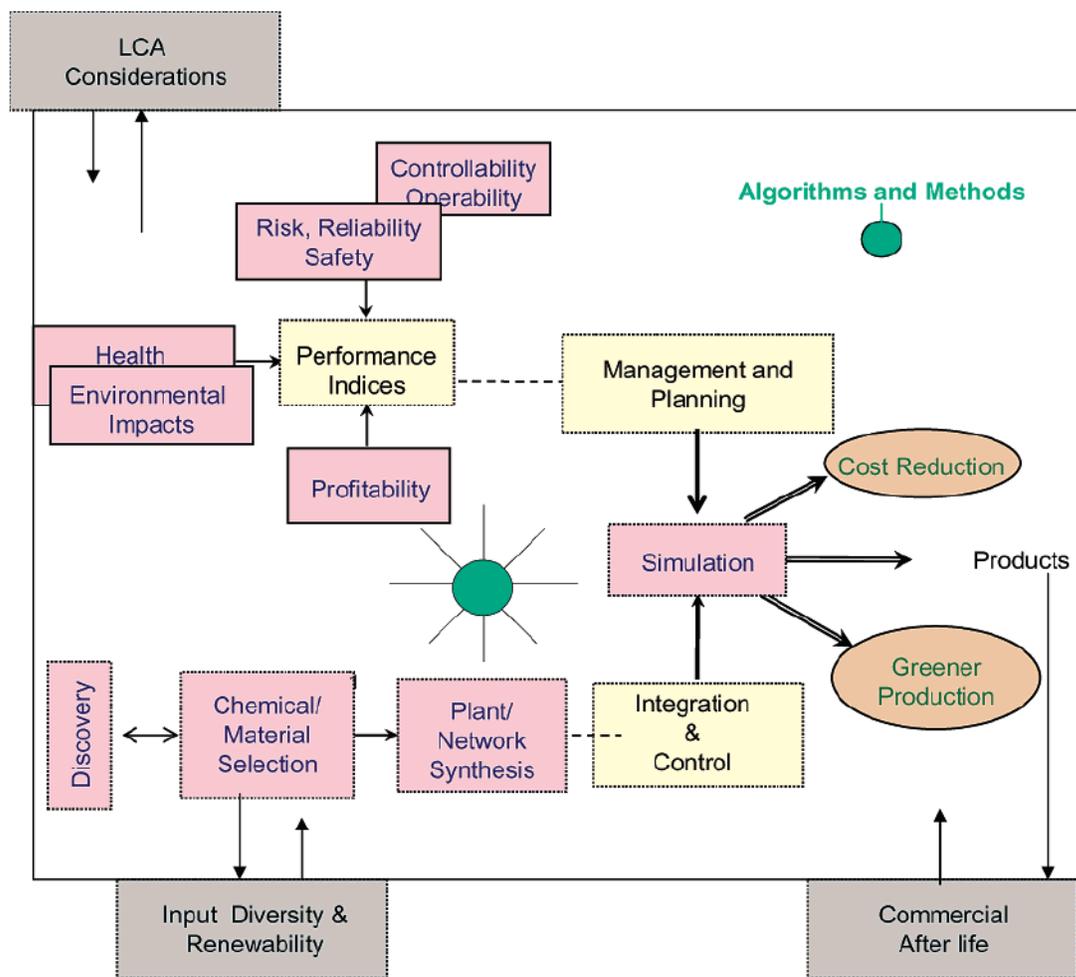


Fig. 1: Integrated framework for environmentally conscious process development and design under uncertainty (2)

Uncertainties are commonly present in chemical and process synthesis due to insufficient experimental data and the lack of accurate models for representing the physical and chemical phenomena. Uncertainties are also encountered over the life cycle of the plant that affect decisions related to plant operations such as process control, production planning and scheduling, supply chain management, reliability and maintenance of the plant.

For example, model uncertainty and external disturbances are important concerns in designing control systems, which are used to minimize deviations from the nominal process conditions and maintaining the safe operation of the plant. Probabilistic approaches and sampling techniques are used commonly to ensure robustness to these model uncertainties. On the other hand, off-line quality control is used to design products and processes that are robust to uncontrollable variation at the design stage. Parameter design strategy is used for this purpose and sampling techniques are employed to propagate the effects of input variability on outputs. The choice of an efficient sampling technique is very important for efficient off-line quality control.

For multipurpose/multiproduct batch plants, optimal production planning and scheduling is important in order to be competitive in a just-in-time production environment. The scheduling problem assigns a sequence of tasks to each equipment over time, according to inventory restrictions and customer demands. The production schedule should be able to accommodate changing product demands, equipment shutdowns and unexpected orders. An extension of the scheduling problem is supply chain management which deals with a complex network of suppliers, plant, warehouses, distribution centers, and customers. Examples of uncertainties in supply chains include fluctuations in product prices, demands or production yields.

In order to increase operational effectiveness and profits, and to save on lost production and costs, chemical plants need to operate with high process reliability and availability. Therefore, reliability issues need to be addressed at the conceptual design stage. Optimal maintenance schedules for the plant need to be determined to increase reliability and availability while maintaining profitability. Uncertainties in equipment availability profoundly affect the profitability of the plant.

Uncertainty analysis and sampling techniques also play an important role in risk assessment and safety. Risk management is a decision making process that is used to reduce the financial and production risk for a business. Environmental risk assessment and financial risk assessment are commonly applied to chemical manufacturing processes. Environmental risk is associated with the toxicity of materials and the effect of hazardous materials on a human population or an entire ecosystem. Financial risk on the other hand, is concerned with pricing decisions and demands. Probability distributions and sampling techniques are frequently used in risk and policy analysis.

The most commonly used sampling technique for uncertainty analysis is Monte Carlo sampling which is based on a pseudo-random number generator. This sampling technique has probabilistic error bounds and large sample sizes are needed to achieve the desired accuracy. Variance reduction techniques have been applied to circumvent the disadvantages of Monte Carlo sampling. Various examples of these variance reduction techniques will be provided in Section 2 (Sampling Techniques). Subsequently, role of sampling in uncertainty analysis and stochastic modeling will be presented (Section 3). Sampling techniques are also important for improving the efficiency of optimization algorithms, which will also be discussed in Section 4. In Section 5, the applications of sampling techniques in product discovery, chemical synthesis, process synthesis and design, process operation and control will be reviewed. Finally in Section 6, conclusions and future trends in sampling techniques will be presented.

## Sampling Techniques

Sampling is a statistical procedure that involves selecting a limited number of observations, states or individuals from a population of interest. A sample is assumed to be representative of the whole population to which it belongs. Instead of evaluating all the members of the population, which would be time-consuming and costly, sampling techniques are used to infer some knowledge about the population.

Sampling techniques could be divided into two groups: *probability sampling* and *non-probability sampling*. In probability sampling, samples are selected based on the theory of probability, which means that each possible set of unit is assigned a probability of selection. The samples are selected by a random process and the confidence intervals for the estimates are known. On the other hand, non-probability sampling does not involve random selection of individuals. An example of this is quota sampling, where the population is first divided into sub-populations and subjects are selected according to judgment or convenience. In this case, the sampling error cannot be determined by probabilistic techniques.

For a good sampling technique, all physically reasonable values of the input and output variables should have some chance of occurring and no region of the population should be excluded. Furthermore, the estimates should be as close as possible to the real values of the quantities being estimated. A good sampling technique also allows an assessment of the relative importance of each input variable.

Probabilistic sampling techniques are based on Monte Carlo methods and are most relevant to this article. They are described in three subsections here: traditional *Monte Carlo Sampling*, *Variance Reduction Techniques* to improve the efficiency of Monte Carlo methods and *Bayesian and Adaptive sampling techniques*, which are used when there is not perfect information about probability. Finally the non-Monte Carlo sampling techniques will be discussed.

### **Monte Carlo Sampling**

One of the simplest and most widely used methods for sampling is the Monte Carlo method. Monte Carlo methods are numerical methods which provide approximate solutions to a variety of physical and mathematical problems by random sampling. The name Monte Carlo, which was suggested by Nicholas Metropolis, takes its name from a city in the Monaco principality, which is famous for its casinos, because of the similarity between statistical experiments and the random nature of the games of chance such as roulette.

Monte Carlo methods were originally developed for the Manhattan Project during World War II to simulate probabilistic problems related random neutron diffusion in fissile material. Although they were limited by the computational tools of that time, they became widely used in many branches of science after the electronic computers were built in 1945. The first publication which presents the Monte Carlo algorithm is probably by Metropolis and Ulam (3).

The basic idea behind Monte Carlo simulation has been that input samples should be randomly generated in order to describe a random output. In crude Monte Carlo approach, a value is drawn at random from the probability distribution for each input, and the corresponding output value is computed. The entire process is repeated  $n$  times producing  $n$  corresponding output values. These output values constitute a random sample from the probability distribution over the output induced by the probability distributions over the inputs. The simplest distribution that is approximated by the Monte Carlo method is a uniform distribution  $U(0,1)$  with  $n$  samples on a  $k$ -dimensional unit hypercube. One advantage of this approach is that the precision of the output distribution may be estimated using standard statistical techniques. On average the error  $\varepsilon$  of approximation is of the order  $O(N^{-1/2})$ . One remarkable feature of this sampling technique is that, the error bound is not dependent on the dimension  $k$ . However, this bound is probabilistic, which means that there is never any guarantee that the expected accuracy will be achieved in a concrete calculation.

The success of a Monte Carlo calculation depends on the choice of an appropriate random sample. The required random numbers and vectors are generated by the computer in a deterministic algorithm. Therefore these numbers are called *pseudorandom numbers* or *pseudorandom vectors*. One of the oldest and best known methods for generating pseudorandom numbers for Monte Carlo sampling is the linear congruential generator (LCG) first introduced by D.H. Lehmer (4). The general formula for a linear congruential generator is the following:

$$I_n = (a \cdot I_{n-1} + c) \bmod m \tag{1}$$

In this formula,  $a$  is the multiplier,  $c$  is the increment which is typically set to zero and  $m$  is the modulus. These are pre-selected constants. The proper choice of these constants is very important for obtaining a sample which performs well in statistical tests. One other pre-selected constant is the *seed*  $I_0$ , which is the first number in the output of a linear congruential generator.

Pseudorandom numbers of different sample sizes on a unit square generated using a linear congruential generator is given in Fig. 2. From this figure it can be seen that pseudorandom number generator produces samples that may be clustered in certain regions of the unit square and does not produce uniform samples. Therefore in order to reach high accuracy, larger sample sizes are needed, which adversely affects the efficiency of this method. Monte Carlo method provides approximate solutions to a variety of mathematical problems. A classic use of Monte Carlo methods is for the evaluation of definite integrals, particularly multidimensional integrals with complicated boundary conditions.

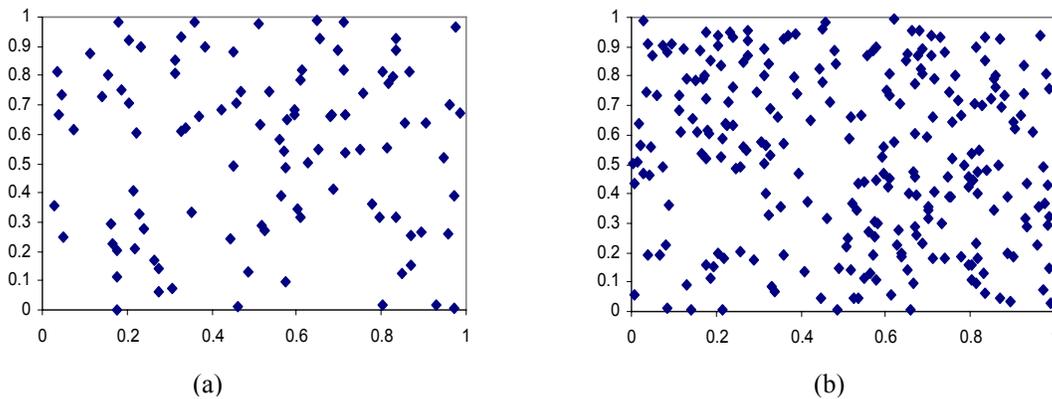


Fig. 2: (a) 100 pseudorandom numbers on a unit square, (b) 250 pseudorandom numbers on a unit square obtained by the linear congruential generator developed by Wichmann and Hill (5)

### Variance Reduction Techniques

For increasing the efficiency of Monte Carlo simulations and overcome the disadvantages such as probabilistic error bounds, variance reduction techniques have been developed.

James (6) has divided variance reduction techniques into four categories. However it should be noted that these categories are not all inclusive. Further, some of the sampling techniques are overlapping between categories.

The first category of sampling techniques extract from a run, more information than immediately evident on the parameter value. An example of this is the *control variate sampling*, which is one of the most versatile variance reduction techniques. Control variates exploit information about the errors in estimates of known quantities to reduce the error in an estimate of an unknown quantity (7). They are best able to estimate the mean of the outcome distribution but also can help in variance estimation. Control variates are set up in order to use a simplified version of a model. One of the problems with

control variates is the selection of effective controls. Also control variates assume a specific probabilistic structure for the simulation output process, usually joint normality of the response and the control variates and this underlying assumption may not always be satisfied (8). These problems restrict the widespread use of control variates. Therefore, it is not surprising that this method is not percolated in process systems engineering.

Sampling techniques in the second category make sure that each individual run is unbiased with respect to the mean outcome measure being estimated. For example in *antithetic sampling*, a negative correlation is introduced between two unbiased estimators of a variable X (9). This technique is applied if there is only one important variable within the model, which is sampled once during a run. Similar to control variate sampling, so far we have not seen any application of this method in chemical engineering literature.

The sampling approaches for variance reduction that are used more frequently for chemical engineering applications are importance sampling, Latin Hypercube Sampling (LHS) (10, 11), Descriptive Sampling (12) and Hammersley Sequence Sampling (HSS) (13, 14). The latter technique belongs to the group of quasi-Monte Carlo methods which were introduced in order to improve the efficiency of Monte Carlo methods by using quasi-random sequences that show better statistical properties and deterministic error bounds. These commonly used sampling techniques are described below with examples.

### ***Importance Sampling***

Importance sampling which may also be called *biased* sampling is a variance reduction technique for increasing the efficiency of Monte Carlo algorithms. Monte Carlo methods are commonly used to integrate a function F over the domain D:

$$\int_D F(x)dx \quad (2)$$

If random numbers are drawn from a normal distribution, information is spread over the interval being sampled. However, if a non-uniform (biased) distribution G(x) is used, which draws more samples from the areas which make a substantial contribution to the integral, the approximation of the integral will be more accurate and the process will be more efficient. This is the basic idea behind importance sampling, where the approximated integral is given by:

$$I = \frac{1}{N} \sum_{i=1}^N \frac{F(x_i)}{G(x_i)} \quad (3)$$

Importance sampling is crucial for sampling low-probability events. The most critical issue for the implementation of importance sampling is the choice of the biased distribution that emphasizes the important regions of the input variables. One of the examples of importance sampling is the Metropolis criterion used in molecular simulations (15). In molecular simulations the configurational phase space is explored and this involves the evaluation of a multi-dimensional integral over 3N degrees of freedom. The crucial feature of the Metropolis approach is that it generates a Markov chain of states and it biases the generation of configurations towards those that make the most significant contribution to the integral. Specifically, it generates states with a probability  $\exp(-\Delta V / k_B T)$ , where  $\Delta V$  is the change in energy,  $k_B$  is the Boltzmann factor and  $T$  is the temperature. This algorithm allows the low energy configurations to be sampled more efficiently, where the Boltzmann factor has an appreciable value. As a result of this thermodynamic properties of fluids could be calculated more accurately.

## ***Stratified Sampling***

Stratification is the grouping of the members of a population into equal or unequal probability areas (strata) before sampling. The strata must be mutually exclusive, which means that every element in the population must be assigned to only one stratum. Also no population element is excluded. It is required that the proportion of each stratum in the sample should be the same as in the population.

*Latin Hypercube Sampling* (LHS) is one form of stratified sampling that can yield more precise estimates of the distribution function (10) and therefore reduce the number of samples required to improve computational efficiency. It is a full stratification of the sampled distribution with a random selection inside each stratum. In Latin hypercube sampling, the range of each uncertain parameter  $X_i$  is subdivided into non-overlapping intervals of equal probability. One value from each interval is selected at random with respect to the probability distribution in the interval. The  $n$  values thus obtained for  $X_1$  are paired in a random manner (i.e., equally likely combinations) with  $n$  values of  $X_2$ . These  $n$  values are then combined with  $n$  values of  $X_3$  to form  $n$ -triplets, and so on, until  $n$   $k$ -tuplets are formed.

Latin Hypercube Sampling (LHS) was designed to improve the uniformity properties of Monte Carlo methods, since it was shown that the error of approximating a distribution by finite sample depends on the equidistribution properties of the sample used for  $U(0,1)$ , and the relationship between successive points in a sample or its randomness or independence is not critical (16).

In *Median Latin Hypercube Sampling* (MLHS), which is a variant of Latin hypercube sampling, the mid-point of the intervals is chosen to sample the uncertain variables. MLHS is similar to the *Descriptive Sampling* described by Saliby (12).

The main drawback of this stratification scheme in LHS and MLHS is that, it is uniform in one dimension and does not provide uniformity properties in  $k$ -dimensions. Sampling based on quadrature (17), cubature techniques (18) or collocation techniques (19) face similar drawback. These sampling techniques perform better for lower dimensional uncertainties. Therefore many of these sampling techniques use correlations to transform the integral into one or two dimensions. However, this transformation is possible only for limited distribution functions when the uncertain variables are tightly correlated. For highly correlated samples similar to what has been observed in thermodynamic phase equilibria, a sampling technique based on confidence region estimates can be used (20).

## ***Quasi-Monte Carlo Methods***

*Quasi-Monte Carlo* methods seek to construct a sequence of points that perform significantly better than Monte Carlo, which has an average case of complexity of the order of  $1/\epsilon^2$ . For a suitably chosen set of samples, the quasi-Monte Carlo method provides a deterministic error bound of the order  $N^{-1}(\log N)^{k-1}$  without any strong assumptions about the integrand. Some well-known quasi-Monte Carlo sequences are *Halton*, *Hammersley*, *Sobol*, *Faure*, *Korobov* and *Neiderreiter* (21). The choice of an appropriate quasi-Monte Carlo sequence is a function of discrepancy. The deterministic upper and lower error bounds of any sequence for integration are expressed in terms of the discrepancy measure. Discrepancy is a quantitative measure for the deviation of the sequence from the uniform distribution. Therefore it is desirable to choose a low-discrepancy sequence. The Halton (22) and Hammersley (23) are some examples of low-discrepancy sequences.

*Hammersley Sequence Sampling* (HSS) is an efficient sampling technique developed by Diwekar and coworkers (13, 14, 24) based on quasi-random numbers. Hammersley Sequence Sampling (HSS) uses Hammersley points to uniformly sample a unit hypercube and inverts these points over the joint cumulative probability distribution to provide a sample set for the variables of interest.

The design of Hammersley points is given below. Any integer  $n$  can be written in radix- $R$  notation ( $R$  is an integer) as follows:

$$n \equiv n_m n_{m-1} \dots n_2 n_1 n_0 \quad (4)$$

$$n = n_0 + n_1R + n_2R^2 + \dots + n_mR^m \quad (5)$$

where  $m = \lceil \log_R n \rceil = \lceil \ln n / \ln R \rceil$  (the square brackets denote the integral part). A unique fraction between 0 and 1 called the inverse radix number can be constructed by reversing the order of the digits of  $n$  around the decimal point as follows:

$$\varphi_R(n) = n_0n_1n_2\dots n_m = n_0R^{-1} + n_1R^{-2} + \dots + n_mR^{-m-1} \quad (6)$$

The Hammersley points on a  $k$ -dimensional cube are given by the following sequence:

$$\vec{z}_k(n) = \left( \frac{n}{N}, \varphi_{R_1}(n), \varphi_{R_2}(n), \dots, \varphi_{R_{k-1}}(n) \right), \quad n = 1, 2, \dots, N \quad (7)$$

where  $R_1, R_2, \dots, R_{k-1}$  are the first  $k-1$  prime numbers. The Hammersley points are  $\vec{x}_k(n) = 1 - \vec{z}_k(n)$ .

In,

Fig. 3, Hammersley points on a unit square are shown. Hammersley Sequence sampling (HSS) technique uses an optimal design scheme for placing  $n$  points on a  $k$ -dimensional hypercube. This scheme ensures that it is more representative of the population showing uniformity properties in multi dimensions, unlike Monte Carlo, Latin Hypercube and its variant Median Latin Hypercube sampling techniques. A qualitative picture of the uniformity properties of the different sampling techniques on a unit square is presented in Fig. 4. It is clearly observed that HSS shows better uniformity than other stratified sampling techniques such as LHS, which are uniform along a single dimension only and do not guarantee a homogeneous distribution of points over the multivariate probability space.

One of the main advantages of Monte Carlo methods is that the number of samples required to obtain a given accuracy of estimates does not scale exponentially with the number of uncertain variables. HSS preserves this property of Monte Carlo. For correlated samples, the approach used by Kalagnanam and Diwekar (13) uses rank correlations (11) to preserve stratified design along each dimension. Although this approach preserves the uniformity properties of the stratified schemes, the optimal location of the Hammersley points is perturbed by imposing the correlation structure. Fig. 5 illustrates the effect of imposing a correlation structure on the sample sets. HSS technique has better performance than LHS and crude Monte Carlo sampling techniques and is at least 3 to 100 times faster for convergence (13).

A variant of the HSS sampling technique is the Latin hypercube Hammersley Sampling (LHSS) (25). The aim of this sampling technique is to better utilize the 1-dimensional uniformity property of LHS and multi-dimensional uniformity property of HSS by coupling them. One dimensional uniformity analysis for Monte Carlo sampling, HSS and LHSS is shown in Fig. 6.

Other variants of Hammersley Sequence Sampling (HSS) are Halton sequence sampling or shifted Hammersley where the first variable is shifted and leaped Halton or Hammersley, where some of the cycles of these sequences are eliminated to improve efficiency for higher dimensional problems. (24, 26) As, the number of dimensions increase, the quasi-random sequences lose their uniformity properties. Therefore, to increase their performance, different quasi-random sequences could be combined and leaping procedure could be applied.

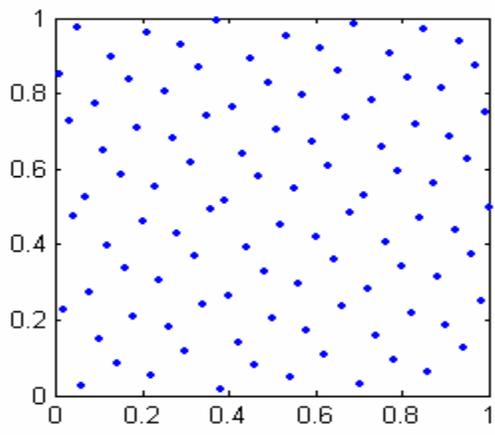


Fig. 3: Hammersley points on a unit square

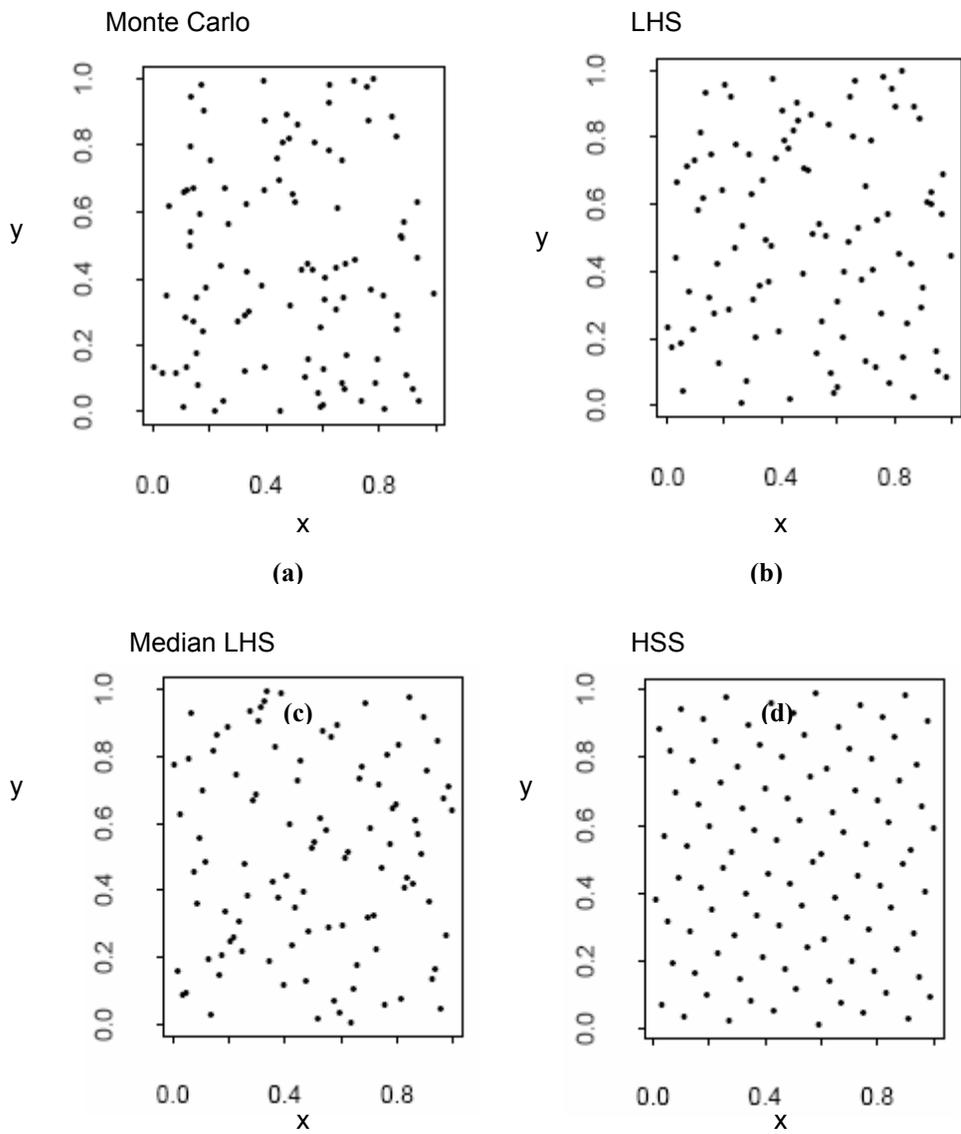


Fig. 4: 100 sample points on a unit square by (a) Monte Carlo sampling (b) Latin hypercube sampling (LHS), (c) Median Latin hypercube sampling (MLHS), (d) Hammersley Sequence Sampling (HSS)

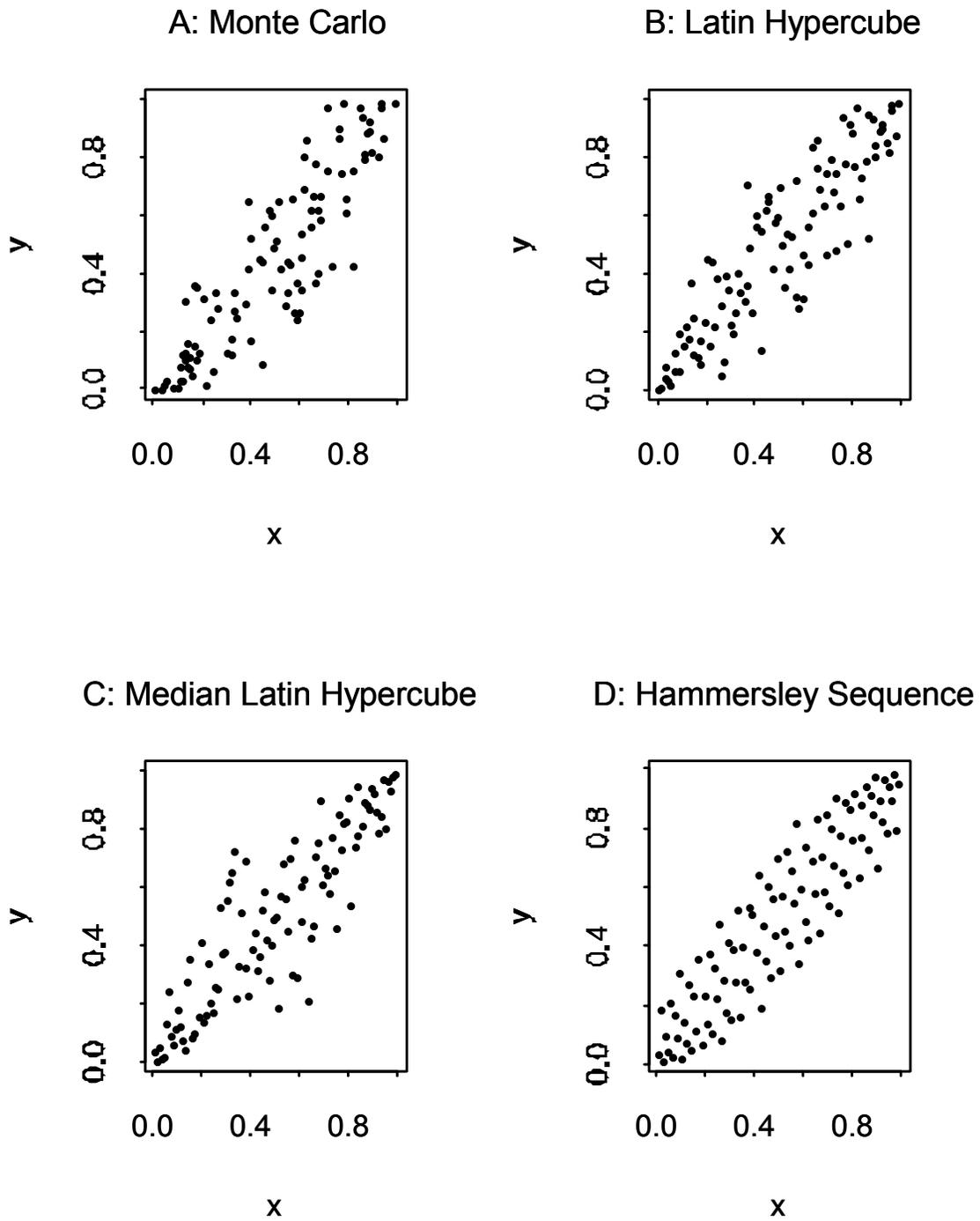


Fig. 5: Sample points (100) on a unit square with correlation 0.9 using (a) Monte Carlo sampling (b) Latin hypercube sampling (LHS), (c) Median Latin hypercube sampling (MLHS), (d) Hammersley Sequence Sampling (HSS)

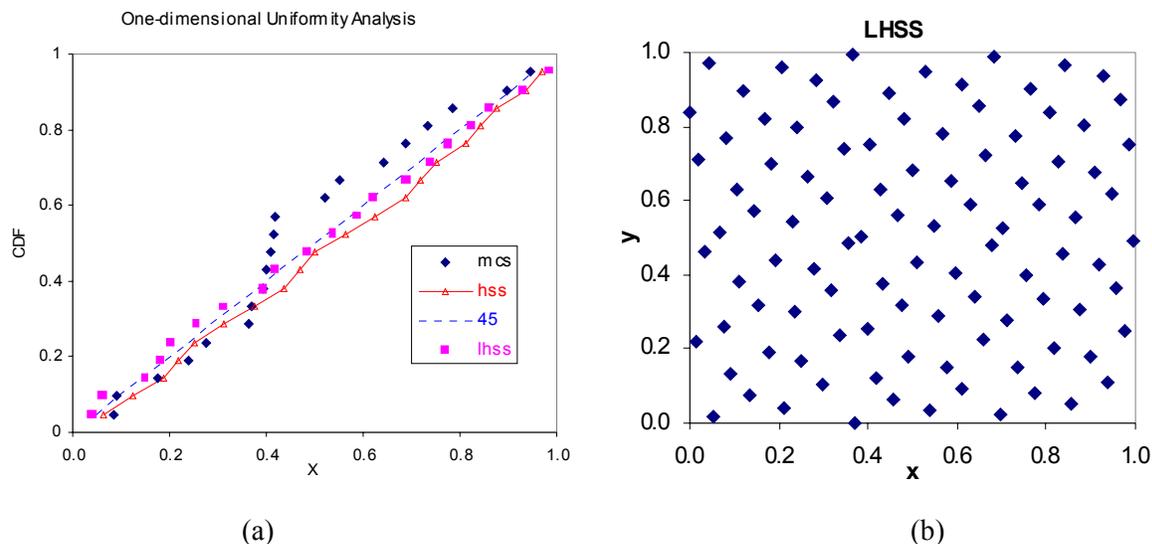


Fig. 6: (a) 1-D uniformity of sample points (20) on a unit square using MCS, HSS and LHSS; (b) 100 sample points on a unit square generated by LHSS (25)

### ***Parallelization of Monte Carlo and Quasi-Monte Carlo Methods***

Monte Carlo and quasi-Monte Carlo sampling techniques discussed above are used to solve a variety of problems in computational chemistry as well as finance or economics where complex models are used. Therefore parallel computing to speed up calculations for these complex models is essential and these sampling techniques should be implemented in a parallel computer architecture, where independent simulations can be performed on different processors.

For reasons of efficiency the random numbers generated for a parallel Monte Carlo simulation using different processors should be un-correlated and they should be generated independently. There are two basic parallelization techniques for generating random numbers. The first method assigns different random number generators for different processors and the second method assigns different substreams of one large random number generator to different processors. When the first method is used, it is possible that there are unknown correlations between the different random number generators in use. Alternatively, if the same random number generator with different parameters is used, one could also encounter similar problems. Hellekalek (27) addressed these issues related to random number generators for a parallel computer architecture with examples.

There are two variations of the second method. The first approach is a “leap-frog” technique, where a substream ( $x_{nL+j}$ ) of lag  $L$  of the original sequence is assigned to the  $j$ -th processor where  $0 < j < L-1$ . The second approach is a “splitting” technique, where the original sequence is partitioned into  $L$  consecutive blocks. Each of the processors is assigned a different block and each block is defined by a unique seed. Both of these methods should be used with caution when the number of dimensions or the sample size is increased.

Mascagni provided a review of the parametrized versions of the pseudo-random number generators for parallel Monte Carlo applications such as linear congruential generators, linear matrix generators, shift-register generators, lagged Fibonacci generators and inversive congruential generators (28).

Quasi-random sequences have also been used in parallel computing in recent years. In order to use quasi-random numbers in parallel, one can break up a single quasi-random number sequence into non-overlapping blocks to be used in parallel processing elements. Comparison of parallel pseudo-random numbers and Sobol sequences has shown that the same kind of accuracy is achieved with the use of a quasi-random sequence in parallel. However, quasi-random sequence (where a block based parallelization is used) converges to the same result in considerably less amount of time (29). Schmid and Uhl (30) also studied the parallelization of quasi-random sequences called the (t,s)-sequences. They have concluded that the block-based parallelization performs much better compared to leaping parallelization for numerical integration. In leaping parallelization, each processing element skips those points handled by other processing elements (leap-frogging).

### ***Bayesian and Adaptive Methods***

Bayesian probability theory was originally developed by Bayes (31). Bayesian and adaptive methods are used when the probability functions are not very accurate. Bayesian method uses two steps. The first step is to identify the conceptual models and the distribution of model parameters. In the second step, the model results are compared with existing observations through a structured probabilistic methodology. In the Bayesian context, a probability represents a degree-of-belief based on all the relevant information at hand. Classical statistical approaches are not very effective in predicting low-frequency, rare but consequential (e.g. accidents or chemical spills) events. Bayesian theory could be applied to these cases (32). A Bayesian approach is also used for sensor fault detection (33).

One of the most important applications of Bayesian methods is to use Bayesian inference for model parameter estimation. This method uses the prior information about the parameters and the likelihood function to find the mode of the posterior distribution (34).

### ***Bayesian Inference***

Bayes theorem (31) states that the posterior probability distribution for an event is proportional to the prior distribution (knowledge) times the likelihood. If we denote  $D$  as the observed data and  $\theta$  as the model parameters, we can write:

$$P(D, \theta) = P(D | \theta)P(\theta) \quad (8)$$

In this equation  $P(D, \theta)$  is the joint probability distribution over all random quantities. This distribution is composed of two parts: a prior distribution  $P(\theta)$  and a likelihood  $P(D|\theta)$ . In order to find the distribution of  $\theta$  conditional on  $D$ , the Bayes theorem is used:

$$P(\theta | D) = \frac{P(\theta)P(D | \theta)}{\int P(\theta)P(D | \theta)d\theta} \quad (9)$$

This is called the posterior distribution of  $\theta$ . The posterior expectation of a function  $f(\theta)$  is:

$$E[f(\theta) | D] = \frac{\int f(\theta)P(\theta)P(D | \theta)d\theta}{\int P(\theta)P(D | \theta)d\theta} \quad (10)$$

It is very difficult to integrate this expression and find  $E[f(\theta)|D]$  especially in high dimensions since for most applications the analytical solution is not available. One of the numerical approaches that have been used is the Markov Chain Monte Carlo (MCMC) method which is described below.

### **Markov Chain Monte Carlo Method**

Let  $X$  be a vector of  $k$  random variables with a distribution  $\pi(\cdot)$ . In Bayesian inference,  $\pi(\cdot)$  will be a posterior distribution. Then the task is to evaluate expressions of the form:

$$E[f(X)] = \frac{\int f(x)\pi(x)dx}{\int \pi(x)dx} \quad (11)$$

A Monte Carlo integration evaluates  $E[f(X)]$  by drawing samples  $\{X_t, t=1, \dots, n\}$  from  $\pi(\cdot)$  and approximates the integral by:

$$E[f(X)] \approx \frac{1}{n} \sum_{t=1}^n f(X_t) \quad (12)$$

The sample averages are used to approximate the expectations. Markov chain Monte Carlo draws these samples by running a cleverly constructed Markov chain for a long time. A Markov chain can be defined as follows:

Suppose a sequence of random variables  $\{X_0, X_1, X_2, \dots\}$  is generated such that at each time step, the next state is sampled from a distribution  $P(X_{t+1}|X_t)$  which depends only on the current state of the chain  $X_t$ . This sequence is called a Markov chain and  $P(X_{t+1}|X_t)$  is called the transition kernel of the chain. One of the transition kernels or updating schemes used in MCMC is the Gibbs transition kernel (35) which is a special case of the general framework of Metropolis and co-workers (36) and Hastings (37).

Many important implementation issues need to be considered for MCMC methods. These include the transition mechanism for the chain, the number of chains to be run and their length and the choice of starting values. These issues are discussed in a review by Brooks (38). MCMC methods and their implementation were also discussed in a textbook by Gilks and co-workers (39). There is also a software called BUGS (Bayesian inference Using Gibbs Sampling) available at the World Wide Web (<http://www.mrc-bsu.cam.ac.uk/bugs/>) for analysis of complex statistical models using MCMC methods.

### **Other Sampling Techniques**

Some other examples of non-Monte Carlo sampling techniques include, systematic sampling, cluster sampling, quota sampling and multi-stage sampling. These sampling techniques are briefly described below.

#### *Systematic sampling*

Systematic sampling is the selection of every  $n^{\text{th}}$  element from a sampling frame. This sampling technique is also called the interval sampling, which means that there is a gap or interval between each selection. This technique is used in industry for quality control where a manufacturer might want to test an item from a production line at certain time intervals to make sure that it satisfies the product specifications, and the equipments and machines are working properly. A random starting point is selected and the sampling interval is chosen in a way that does not create a pattern that would threaten randomness. More information about systematic sampling can be found in Madow and Madow (40).

#### *Cluster sampling*

In cluster sampling, the entire population is divided into clusters, or groups and a random sample is selected from these clusters. When the researcher does not have enough information about

the individual members of a population but can get a complete list of the groups or clusters, this sampling technique would be useful (41, 42, 43).

#### *Quota sampling*

In quota sampling, the population is first divided into mutually exclusive sub-populations, just as in *stratified sampling*. Then the subjects are selected according to judgment or easy availability from each sub-population. This sampling technique is often used in opinion polling and market research. This is not a random sample; therefore, statistical methods cannot be applied to measure the sampling error. A discussion on the validity of inferences made from quota sampling was presented by Smith (44).

#### *Multi-stage sampling*

This sampling technique involves the selection of a sample in at least two stages. In the first stage large groups of clusters are selected and in the second stage population units are selected from the clusters to derive a final sample. For an application of multi-stage sampling in epidemiology please refer to refs 45 and 46).

These sampling techniques are mostly designed for convenience and efficiency. However they are not as accurate and many times the sampling error cannot be estimated by classical statistical techniques. They are more frequently used in areas such as market research, polling, or interviewing to infer some knowledge about a population. They are also encountered in epidemiology studies, where the causes and prevalence for certain diseases are statistically analyzed. However, they do not have a wide applicability in process systems engineering.

## **Uncertainty Analysis and Stochastic Modeling**

Role of sampling in uncertainty analysis is indisputable and encompasses all application areas in process design, operation and control. The uncertainties commonly encountered in chemical systems can be divided into two groups (47): (1) static uncertainties; (2) dynamic uncertainties.

#### ***Static Uncertainties***

Static uncertainties are represented by probability distributions. Inclusion of uncertainties in a deterministic model results in a stochastic model. Stochastic modeling is an iterative procedure which consists of these four steps (48) as shown in Fig. 7.

1. Uncertainty quantification which involves specifying uncertainties in key input parameters in terms of probability distributions
2. Sampling distribution of the specified parameter in an iterative fashion
3. Propagating the effects of uncertainties through the model
4. Applying statistical techniques to analyze the results.

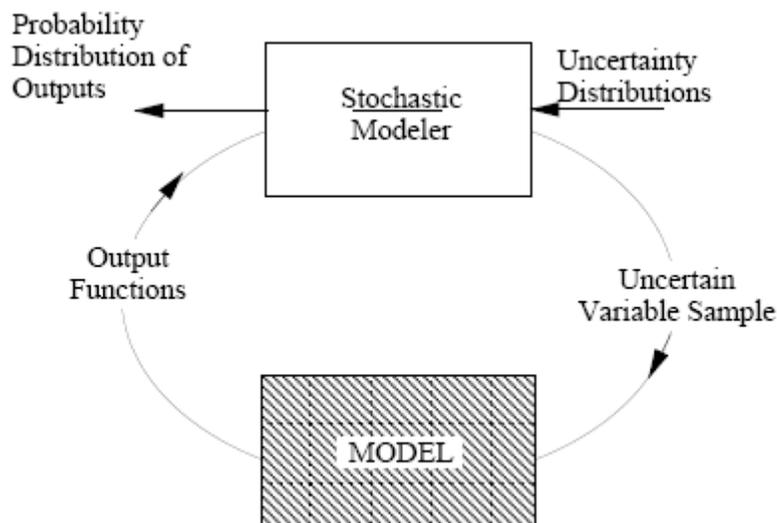


Fig. 7: Stochastic modeling framework

In the first step of stochastic modeling framework, uncertainties in key input variables are represented by probability distribution functions. An example of uncertainty characterization and quantification by probability distributions was presented by Kim and Diwekar (49) in a computer-aided molecular design (CAMD) problem. Discrepancies between the experimental data for predicting a thermodynamic property and the models are commonly encountered in CAMD. For example, Fig. 8 shows the uncertainties in more than 1800 interaction parameters present in the UNIFAC activity coefficient model to predict solvent selection objectives for acetic acid separations. Uncertainty factors (UFs) were established as the ratio between the experimental and the calculated values of activity coefficients at infinite dilution  $\gamma^\infty$  as defined in equation 13. Furthermore, uncertainty factors were divided into three categories based on the type of family: organic/water (lognormal distribution), water/organic (normal distribution) and organic/organic(lognormal distribution).

$$UF = \frac{\gamma_{exp}^\infty}{\gamma_{calc}^\infty} \quad (13)$$

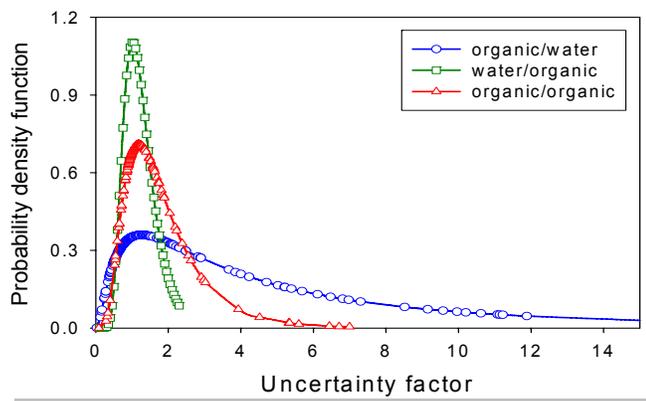


Fig. 8: Probability density functions for uncertainty factors for organic/water, water/organic, organic/organic families (50)

The type of distribution for an uncertain variable is a function of amount data available and the characteristic of the distribution function. The simplest distribution for an uncertain variable is a uniform distribution, which has a constant probability. This means that the uncertain variable can take any value within an interval  $[a,b]$  with equal probability. On the other hand, if the uncertain variable is represented by a normal (Gaussian) distribution, there is a symmetric but equal probability that the value of the uncertain variable will be above or below a mean value. In log-normal or some triangular distributions, there is a higher probability that the value of uncertain variable will be on one side of the median, resulting in a skewed shape. A beta distribution provides a wide range of shapes and is a very flexible means of representing variability over a fixed range. In some special cases, user-supplied distributions are used such as chance distribution.

Once probability distributions are assigned to uncertain parameters, the next step is to perform a sampling operation from the uncertain parameter domain. Then, the uncertainties are propagated through the model. The stochastic modeler assigns the specified distributions to the input parameters and using sampling methods described in the previous section (e.g. Monte Carlo, LHS, HSS), the sampled values of each uncertain variable are passed through the model. After a model 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. After all samples or observations have gone through the cycle for a specified number of times (typically 20 to 100 or more, depending on the accuracy sought by the user), the outputs are collected in terms of cumulative probability density functions.

### ***Dynamic Uncertainties***

Dynamic uncertainties are also ubiquitous in chemical systems, especially for batch processes. Due to the dynamic nature of these processes, even some of the static uncertainties are translated into dynamic uncertainties. An example of this is shown in Fig. 9. This figure shows how the uncertainties in activity coefficients predicted by the UNIFAC method affect the time-dependent relative volatility profile in a batch distillation column (47). Despite this fact, a generalized method of treatment for dynamic uncertainties for chemical systems was presented only recently (47, 51, 52). This method is based on Ito processes and real options theory from finance literature.

Ito processes are a large class of continuous time stochastic processes. One of the simplest examples of a stochastic process is the random walk process. The Wiener process, also called a Brownian motion, is a continuous limit of the random walk and serves as a building block for Ito processes, through the use of proper transformations.

A Wiener process satisfies three important properties. Firstly, it satisfies the Markov property. The probability distribution for all future values of the process depends only on its current value. Secondly, it has independent increments. The probability distribution for the change in the process over any time interval is independent of any other time interval (non-overlapping) and thirdly, changes in the process over any finite interval of time are normally distributed, with a variance which is linearly dependent on the length of time interval,  $dt$ . The general equation of an Ito process is given below:

$$dx = a(x,t)dt + b(x,t)dz \quad (14)$$

In this equation,  $dz$  is the increment of a Wiener process, and  $a(x,t)$  and  $b(x,t)$  are known functions. There are different forms of  $a(x,t)$  and  $b(x,t)$  for various Ito processes. In this equation,  $dz$  can be expressed by  $dz = \varepsilon_t \sqrt{dt}$ , where  $\varepsilon_t$  is a random number drawn from a unit normal distribution.

The simplest generalization of equation 14 is the equation for Brownian motion with drift given by:

$$dx = \alpha dt + \sigma dz \quad (15)$$

where  $\alpha$  is called the drift parameter, and  $\sigma$  is the variance parameter.

Other examples of Ito processes are the geometric Brownian motion with drift given below in equation 16 and the geometric mean reverting process given in Equation 17. Also it has been shown that the relative volatility profile in Fig. 18(b) can be represented by a geometric mean reverting process (47):

$$dx = \alpha x dt + \sigma x dz \quad (16)$$

$$dx = \eta (\bar{x} - x) dt + \sigma x dz \quad (17)$$

where  $\eta$  is the speed of reversion and  $\bar{x}$  is the nominal level that  $x$  reverts to. In geometric Brownian motion, the percentage changes in  $x$  and  $\Delta x/x$  are normally distributed (absolute changes are lognormally distributed). In geometric mean reverting processes, the variable may fluctuate randomly in the short run, but in the longer run it will be drawn back towards the marginal value of the variable. The expected change in  $x$  depends on the difference between  $x$  and  $\bar{x}$ . If  $x$  is greater (less) than  $\bar{x}$ , it is more likely to fall (rise) in the next short interval of time. The variance also grows with  $x$ .

The dynamic uncertainties in chemical processes (batch processes) could be represented by these Ito processes, depending on the character of uncertainty. For example, the thermodynamic uncertainties in batch processes were modeled using Ito processes and system non-idealities were easily distinguished (47). The parameters of the Ito process are estimated based on a regression analysis technique. For more details on Ito processes please refer to refs. 1, 52 and 53.

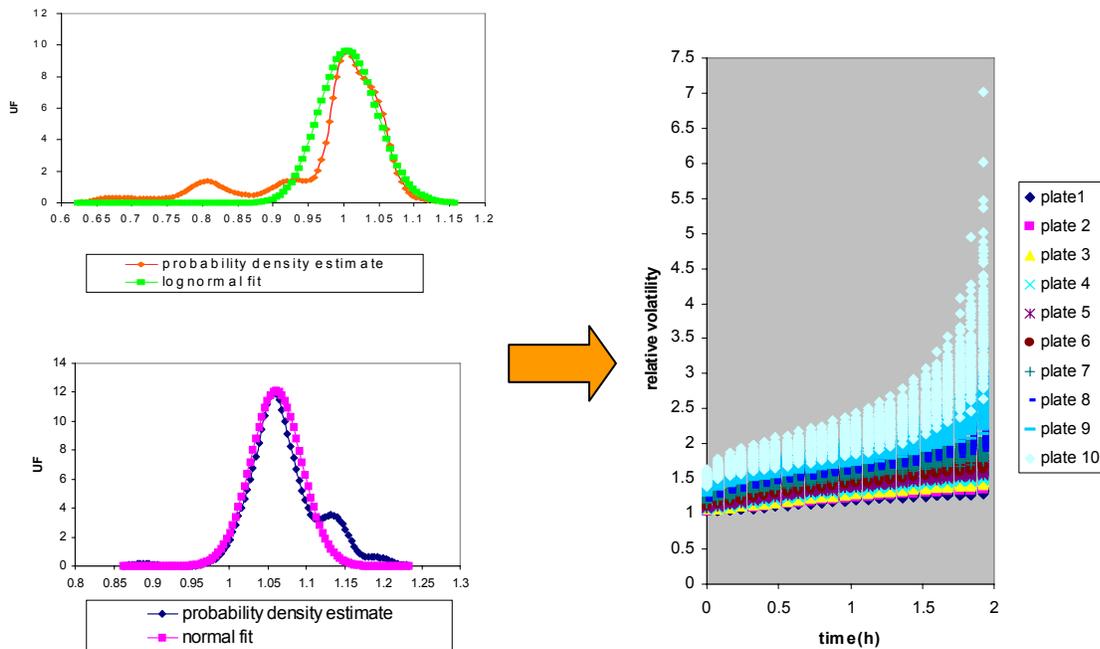


Fig. 9: (a) The uncertainties in activity coefficients predicted by UNIFAC; (b) The effect of static uncertainties on relative volatility profile on each plate in a batch distillation column (47)

## Efficiency Improvements in Optimization Algorithms

The role sampling plays in optimization algorithms extends beyond uncertainty analysis. Sampling accuracy is also crucial for deriving efficient algorithms for discrete optimization and multi-objective optimization problems, which will be described in this section

### *Discrete Optimization*

Discrete optimization problems involve discrete decisions and combinatorics. Discrete optimization problems are classified into groups such as integer programming (IP), mixed integer linear programming (MILP) and mixed integer nonlinear programming (MINLP). Many chemical engineering applications like chemical synthesis, process synthesis, planning and scheduling involve discrete decision variables and mixed integer problems. Probabilistic combinatorial methods can be used to solve these problems. Examples of these methods are simulated annealing (SA) and genetic algorithms (GA). If the solution space is discontinuous or if the systems have large combinatorial explosion, these probabilistic methods provide an alternative to mathematical programming techniques such as branch and bound, Generalized Bender's Decomposition (GBD) and Outer Approximations (OA) traditionally used to solve discrete optimization problems (52).

Simulated annealing (SA) is a heuristic combinatorial optimization method. Simulated annealing utilizes the analogy between the annealing procedure, where a metal cools and freezes into its minimum energy structure and the search for a minimal value in an optimization problem. In SA, the objective function (which is usually the cost) becomes the energy of the system. The goal is to minimize the cost (energy). Random permutations are generated to displace particles, which is analogous to moving the system to another configuration. If the configuration that results from the move has a lower energy state, the move is accepted. Otherwise the move is accepted according to the Metropolis criteria (accepted with a probability =  $\exp(-\Delta E/K_b T)$ ) (54). At high temperatures, a large percentage of uphill moves are accepted. As the temperature gets cooler, a small percentage of uphill moves are accepted. After the system has evolved to thermal equilibrium at a given temperature, the temperature is lowered and the annealing process continues until the system reaches the "freezing" temperature. Painton and Diwekar (55) used the simulated annealing (SA) technique to improve the performance of space nuclear power plants.

As SA is a probabilistic method, several random probability functions are involved in this algorithm. The random probability  $A_{ij}$  is used for acceptance determination in Metropolis criterion, while the random generation probabilities  $G_{ij}$  are used to generate subsequent configurational moves. The  $G_{ij}$  of the conventional SA algorithms rely on pseudo-random number generators such as Monte Carlo sampling, which result in clustered moves over the configurational space. Therefore, a larger number of moves or generations are needed to cover the configurational space more evenly and this results in a longer Markov chain length (i.e. number of moves) at each temperature level. As mentioned earlier, HSS technique can generate quasi-random samples showing k-dimensional uniformity properties. The HSS technique was used to develop a new SA algorithm called the efficient simulated annealing (ESA). Since HSS generates more uniform samples in the multivariate space, it requires less number of moves to approximate ideal probabilities.

Fig. 10 shows the trajectories of the objective value for the test function  $f(y) = \sum_{i=1}^{10} y_i^2$  with different Markov chain lengths. ESA found the solution with a Markov chain length of 45 at each temperature while the traditional SA needed a Markov chain length of 75 to reach the same solution. ESA was found to be approximately 30~54% more efficient than conventional SA

(56).

Genetic algorithms (GA) are also used for combinatorial optimization problems. GA's follow a search procedure based on the Darwin's theory of evolution and the idea of survival of the fittest. GA's begin with a set of solutions (represented by chromosomes) which is called the initial population. The population for the next generation is selected according to a randomized selection procedure involving four operators: (1) reproduction; (2) crossover; (3) mutation and immigration and the fittest individuals are selected from the population for the next generation. This procedure is repeated until a stopping criterion such as number of populations or improvement of the best solution is satisfied. Conventional GAs also use Monte Carlo sampling based on pseudo-random numbers for generating the initial population and various genetic operators. Similar to the efficient simulated annealing algorithm (ESA), an efficient genetic algorithm (EGA) based on the use of Hammersley sequence sampling technique was developed by Diwekar and Xu (57), in order to improve the efficiency of conventional GA's. Table 1 shows the comparison of GA's for problems varying in complexity and size.

Table 1: Efficiency Improvement in GA Using the Hammersley Sequence Sampling (HSS) Technique (47)

problems	number of dimensions (nd)	optimal value	generation		efficiency improvement
			MGA (Monte Carlo)	EGA (HSS)	
problem 1	10	0	15	4	73.33%
	20	0	43	10	76.74%
problem 2	3-11	0	9	6	33.33%
problem 3	5	-1	176	83	52.84%

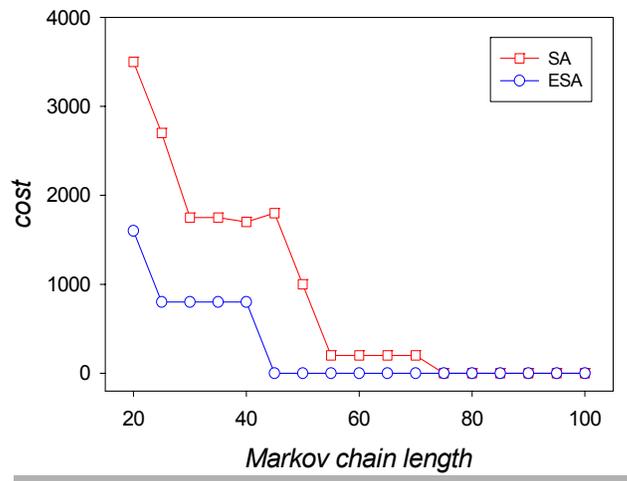


Fig. 10: The comparison between SA and ESA (52)

### Optimization under Uncertainty

Optimization under uncertainty refers to the branch of optimization problems where there are uncertainties involved in the data or the model, and is popularly known as stochastic programming or stochastic optimization problems. The generalized stochastic framework to solve optimization problems under uncertainty involves two recursive loops: (1) the sampling loop and (2) the optimization loop. A schematic representation of this stochastic framework is shown in **Error!**

**Reference source not found.** By interchanging the position of the sampling loop, two kinds of solution procedures could be obtained. These are called “here and now” and “wait and see” problems. “Here and now” problems yield optimal solutions to achieve a given level of confidence. On the other hand, “wait and see” problems involve a category of formulations that show the effect of uncertainty on optimum design. A here and now problem replaces a deterministic model by an iterative stochastic model with sampling loop representing the discretized uncertainty space as shown in **Error! Reference source not found.** A wait and see problem involves deterministic optimal decision at each scenario or random sample, equivalent to solving several deterministic optimization problems.

Both here and now and wait and see problems require representation of uncertainties in the probabilistic space and then propagation of these uncertainties through the model to obtain probabilistic representation of output. Here sample approximation methods and sampling accuracy are often used to derive new algorithms for optimization under uncertainty.

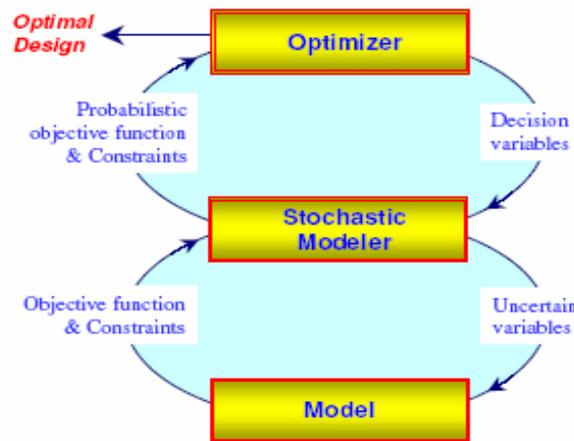


Fig. 11: Stochastic programming framework

### Sample Approximation Methods

As stated earlier, the stochastic programming formulations often include some approximations of the underlying probability distribution. The disadvantage of sampling approaches that solve the  $\gamma^{\text{th}}$  approximation completely is that some effort might be wasted on optimizing when approximation is not accurate (58). For stochastic linear programming the L-shaped method is a commonly used technique (59). For specific structures where the L-shaped method is applicable, two approaches avoid these problems by embedding sampling within another algorithm without complete optimization. These two approaches are the method of Dantzig and Infanger (60) which uses importance sampling to reduce variance in each cut based on a large sample, and the stochastic decomposition method proposed by Higle and Sen (61). For more details on stochastic programming please refer to refs. 1, 52 and 58.

Sample average approximation methods have also been used to reduce the computational time and increase accuracy for stochastic process design problems. Wei and Realff (62) presented a method which involves two algorithms: optimality gap method (OGM) and confidence level method (CLM) to solve convex stochastic mixed integer nonlinear problems. A smaller sample size is used to make

decisions (with several replications) and a larger one is used to re-evaluate the objective value with the decision variables fixed. The sample sizes and replication number are increased until a stopping criterion is satisfied. In the OGM algorithm, the sample sizes are increased until the optimality gap of each upper and lower bound is sufficiently small; in the CLM algorithm, the sample sizes are increased until an overall accuracy probability is within a certain tolerance.

### *Sampling Accuracy and Optimization*

In almost all stochastic optimization problems, the major bottleneck is the computational time involved in generating and evaluating probabilistic functions which represent the objective function and constraints. The number of samples required for a given accuracy in stochastic optimization problem depends upon factors such as type of uncertainty and the point values of the decision variables (55). Especially, for optimization problems, the number of samples required depends on the location of the trial point solution in optimization space. Fig. 12 shows how the shape of the surface over a range of uncertain parameter values changes at different optimization iterations. Therefore, the selection of the number of samples is an important step which ultimately decides the accuracy of the optimal solution in stochastic programming.

For the solution of stochastic integer programming problems, variants of simulated annealing (SA) and genetic algorithms (GA) have been developed.

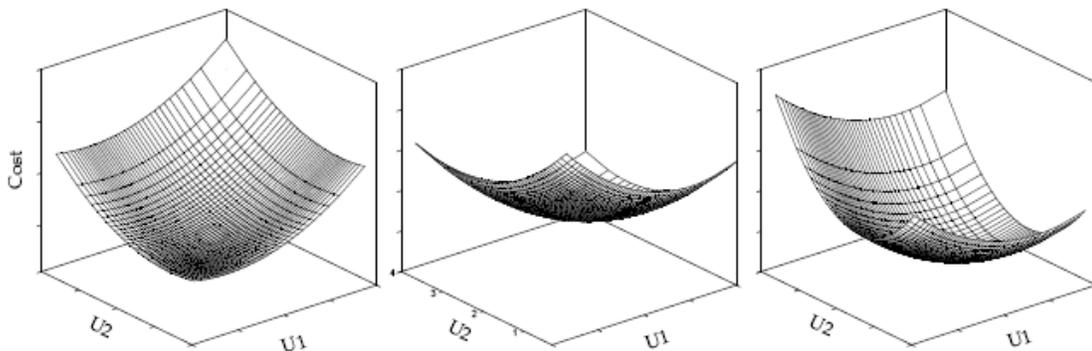


Fig. 12: Uncertainty space at different optimization iterations (Diwekar, 2003a)

### *Stochastic Annealing and Stochastic Genetic Algorithms*

The stochastic annealing (STA) algorithm (55, 63, 64) is a variant of simulated annealing and is used to optimize stochastic integer programming problems. The stochastic annealing provides an improvement over simulated annealing by obtaining both the decision variables and the number of samples required for the optimization problem. For balancing computational efficiency and solution accuracy, a penalty function is introduced in the objective function to ensure that the algorithm selects greater number of samples as the solution nears optimum value.

Annealing temperature schedule (cooling schedule), is used to decide the weight  $b(t)$  on the penalty term for imprecision in the probabilistic objective function. The choice of the penalty term also depends on the error bandwidth ( $\epsilon$ ) of the function that is optimized and must incorporate the effect of number of samples. Therefore, the new objective function in stochastic annealing, consists of a probabilistic objective value  $P$  and the penalty function,  $(b(t)\epsilon)$ :

$$\min Z(\text{cost}) = P(x;u) + b(t)\varepsilon \quad (18)$$

The weighting function  $b(t)$  can be expressed in terms of the temperature levels ( $t$ ) and is given by  $b(t)=b_0/k^t$  where  $b_0$  and  $k$  are constants. At high temperatures, the sample size is small, and the algorithm is mainly exploring the functional topology to identify regions of optima. As the system gets cooler, the algorithm searches for the global minimum and more accurate estimates of the objectives/costs are needed and this requires more samples. The error bandwidth of the Monte Carlo samples ( $\varepsilon_{\text{MCS}}$ ) is estimated by the central limit theorem.

Using the Hammersley Sequence Sampling (HSS) technique for the generation probability  $G_{ij}$ , Kim and Diwekar (56) developed the efficient stochastic annealing algorithm (ESTA). This algorithm uses the central limit theorem to evaluate the sampling errors and uses the HSS technique for generation probability. Another variant of STA is the Hammersley stochastic annealing algorithm (HSTA). This algorithm was presented by Kim and Diwekar (56) and it uses (1) HSS for the generation probability  $G_{ij}$  in annealing procedure; (2) HSS for the inner sampling loop where number of samples  $N_{\text{samp}}$  are determined, and (3) the HSS specific error bandwidth ( $\varepsilon_{\text{HSS}}$ ). The error bandwidth for the Hammersley sequence samples ( $\varepsilon_{\text{HSS}}$ ) is given by a fractal dimension analysis (64, 65). This methodology uses the  $k$ -dimensional uniformity properties of HSS technique and the HSS error bandwidth to achieve a trade-off between accuracy and efficiency. The HSTA algorithm is a useful tool for solving large scale combinatorial optimization problems under uncertainty and was applied to computer aided molecular design problems (56).

A similar approach was also applied to genetic algorithms by Diwekar and Xu (57). First the stochastic genetic algorithm (SGA) was developed which employs Monte Carlo sampling for stochastic optimization problems. Then efficient stochastic genetic algorithm was developed which uses HSS technique and Monte Carlo confidence intervals. Finally, the Hammersley stochastic genetic algorithm (HSGA) was introduced which uses HSS technique and HSS specific error bandwidth to achieve a trade-off between accuracy and efficiency. HSGA displayed the best performance among these algorithms. The performance of these three algorithms is compared in Fig. 13 for the test function:

$$f(x, y, \zeta) = \sum_i^{ND} \left( \zeta_i x_i - \frac{i}{ND} \right)^2 + \sum_i^{ND} \zeta_i y_i^2 - \prod_i^{ND} \cos(4\pi \zeta_i y_i) \quad (19)$$

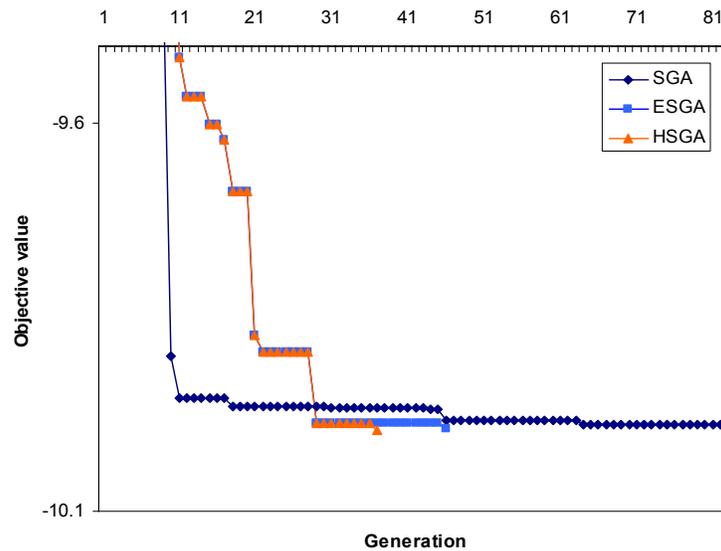


Fig. 13: Comparison of performance and convergence path for SGA, ESGA and HSGA

### **Multi-Objective Optimization**

Multi-objective optimization (MOP) problems deal with conflicting and different objectives. They are commonly used where engineers are not only looking for low cost options but also trying to reduce the environmental and health impacts and risk and improve the reliability and safety of the plant. A generalized multi-objective optimization problem can be formulated as follows:

$$\begin{aligned} \min \quad & Z = Z_i, i = 1, \dots, p \quad p \geq 2 \\ \text{s.t.} \quad & h(x, y) = 0 \\ & g(x, y) \leq 0 \end{aligned} \tag{20}$$

where  $x$  and  $y$  are continuous and discrete decision variables, and  $p$  is the number of objective functions. The functions  $h(x, y)$  and  $g(x, y)$  represent equality and inequality constraints, respectively. There is a large array of analytical techniques to solve this MOP problem; however, the MOP methods are generally divided into two basic types: preference-based and generating methods. Preference-based methods such as goal programming attempt to quantify the decision-maker's preference, and with this information, the solution that best satisfies the decision-maker's preference is then identified (2, 52). As is well-known, mathematics cannot isolate a unique optimum when there are multiple competing objectives. Mathematics can at most aid designers to eliminate design alternatives dominated by others, leaving a number of alternatives in what is called the Pareto set (66). Generating methods, such as the weighting method and the constraint method, have been developed to find the exact Pareto set or an approximation of it. For each of these designs, it is impossible to improve one objective without sacrificing the value of another relative to some other design alternative in the set. From among the dominating solutions, it is then the design which is the most appropriate for that particular purpose is selected. At issue is an effective means to find the members of the Pareto set for a design problem, especially when there are more than two or three objectives; the analysis per design requires significant computations to complete, and there are an almost uncountable number of design alternatives. A pure algorithmic approach to solving is to select one to minimize while the remaining objectives are turned into an inequality constraint with a parametric righthand side,  $L_k$ . The problem takes on the following form:

$$\begin{aligned} \min \quad & Z = Z_j \\ \text{s.t.} \quad & h(x, y) = 0 \\ & g(x, y) \leq 0 \\ & Z_k \leq L_k, k = 1, \dots, j-1, j+1, \dots, p \end{aligned} \tag{21}$$

where  $Z_j$  is the chosen  $j^{\text{th}}$  objective that is to be optimized. Solving repeatedly for different values of  $L_k$  chosen between the upper,  $Z_U(j)$  and the lower,  $Z_L(j)$ , bounds leads to the Pareto set. This is the basis of the MINSOOP algorithm.

MINSOOP (Minimizing Number of Single Objective Optimization Problems) algorithm was developed by Fu and Diwekar (67) to address multi-objective optimization problems based on Hammersley Sequence Sampling (HSS) technique. The steps for a multi-objective optimization problem with  $k$  objectives (to be minimized) are listed as follows:

Step 1: Solve  $k$  single objective optimization problems individually with the original constraints of a multiobjective problem to find the optimal solution for the individual  $k$  objectives.

Step 2: Compute the value of each of the  $k$  objectives at each of the  $k$  individual optimal solutions. In

this way, an approximation of the potential range of values for each of the  $k$  objectives is determined and saved in a table (called payoff table). The minimum possible value is the individual optimal (minimizing) solution. The approximate maximum possible value of the Pareto set is the maximum value for that objective found when minimizing the other  $k-1$  objectives individually.

Step 3: Select a single objective ( $Z_l$ ) to be minimized. Transform the remaining  $k-1$  objectives into equality constraints of the form  $Z_i \leq \varepsilon_i$ ,  $i=1, \dots, k$ ,  $i \neq l$  and add these new  $k-1$  constraints to the original set of constraints. Then the original multi-objective optimization problem is transformed into a family of single objective optimization problems with parametric right hand sides.

Step 4: Select a desired number of single objective optimization problems to be solved to represent the Pareto set. Using the HSS technique to generate the desired number of combinations of the inequality constraint values  $\varepsilon_1, \dots, \varepsilon_{l-1}, \varepsilon_{l+1}, \dots, \varepsilon_k$  within the range determined in step 2.

Step 5: Solve the constrained problems set up in step 4 for every combination of the right hand side values determined in step 3. These feasible solutions form an approximation for the Pareto set.

The uniformity property of HSS technique is crucial for the success of MINSOOP algorithm. Fig. 14 shows how the MINSOOP algorithm improves efficiency for a simple nonlinear convex optimization problem as the number of objectives increases. Similar improvements are noted in multi-objective genetic algorithms based on the uniformity property of HSS (68).

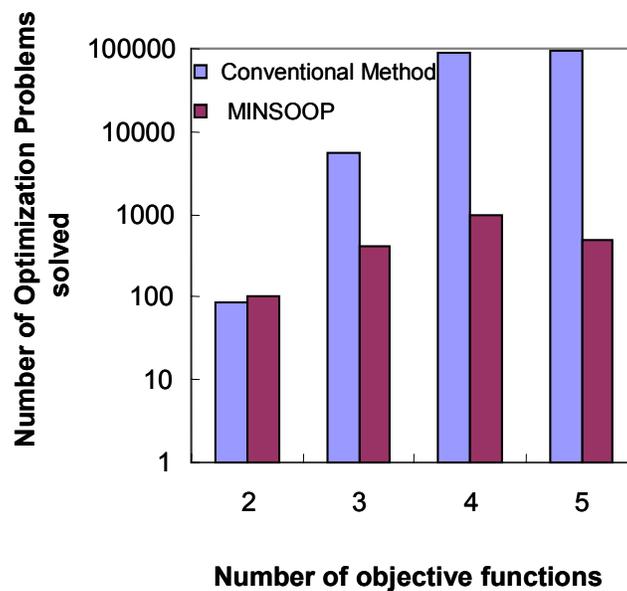


Fig. 14: Computational speed-up through MINSOOP (Fu and Diwekar, 2004)

## Application of Sampling Techniques in Chemical Systems

The life cycle of a chemical manufacturing process extends from product discovery, raw material selection (chemical synthesis) and process development (process synthesis and design) to process operation, planning and management which involve tasks such as scheduling, supply chain, and process control. Contemporary process design approaches require engineers to not only look for low cost options but also include several other criteria such as reliability, flexibility, operability, controllability, environmental and ecological impacts, safety and quality into different stages of analysis and design. This results in additional complexities and uncertainties. This section will present some of the applications of sampling techniques in product discovery, chemical synthesis, process synthesis and process operations ranging from scheduling, supply chain and process control. The role of sampling techniques in risk and reliability analysis will also be discussed.

### *Product Discovery and Design; Computational Chemistry and Molecular Simulations*

At the discovery stage of a chemical process, computational chemistry, molecular modeling and simulations are widely used in chemical and pharmaceutical industries. For the design and discovery of new molecules or drug compounds, Monte Carlo or molecular dynamics simulations are applied to estimate the physical, chemical, biological and toxicological properties of interest. A review of molecular modeling and simulation techniques can be found in refs. 15 and 69. These methods depend heavily on random number generators and sampling. The following discusses the role of sampling in molecular simulations where Monte Carlo methods are predominant.

In Monte Carlo molecular simulations, particles are randomly selected and moved by a random extent and the energy change of the system is analyzed. These random perturbations on the system configuration are accepted according to Metropolis criterion, with a probability proportional to the Boltzmann constant. This forms the basis of Metropolis Monte Carlo (MMC) method which is the first approach in increasing efficiency of Monte Carlo molecular simulations by using the “importance sampling” concept. In importance sampling, a biased distribution is used to obtain more samples from a region of importance. Boltzmann distribution function is such a distribution used in MMC where system configuration states which make substantial contributions to the ensemble averages are generated. The Metropolis Monte Carlo method requires large number of samples to generate accurate property estimations and is computationally intensive especially for large number of molecules and complex fluids. Therefore many researchers have worked on sampling techniques in order to speed up the calculations and cover the configurational space more efficiently. Some examples of these biased sampling techniques are configurational-bias Gibbs ensemble (70) and its variants, (71-73) and non-Boltzmann biasing techniques (74, 75). Despite the fact that, the literature is abundant in examples of algorithms for efficiency improvement for Monte Carlo simulations, most of these methods are derived based on the importance sampling (biased sampling) principle and are often problem specific and offer customized solutions for particular systems.

Recently, a universal approach for increasing efficiency in molecular simulations using the Hammersley sequence sampling (HSS) technique was presented (76). Pseudo-random numbers are used in Metropolis Monte Carlo method for performing the random moves for the molecules and for acceptance probability. This new technique replaces the pseudo-random numbers in a systematic way by quasi-random samples of HSS, to speed-up the simulations and to increase the accuracy. While replacing the pseudo-random numbers with quasi-random samples, the k-dimensional uniformity property of HSS technique was maintained and exploited to cover the configurational space more efficiently. This method was used to estimate thermodynamic and biological/toxicological properties of chemicals, more specifically octanol-water partition coefficients. This new framework provided a

better way to predict octanol-water partition coefficients in terms of prediction accuracy and computational efficiency (number of cycles) as shown in Fig. 15. It should also be noted that this proposed approach can be used in conjunction with the biased MCS (importance sampling) strategies presented in literature and is not restricted to specific applications.

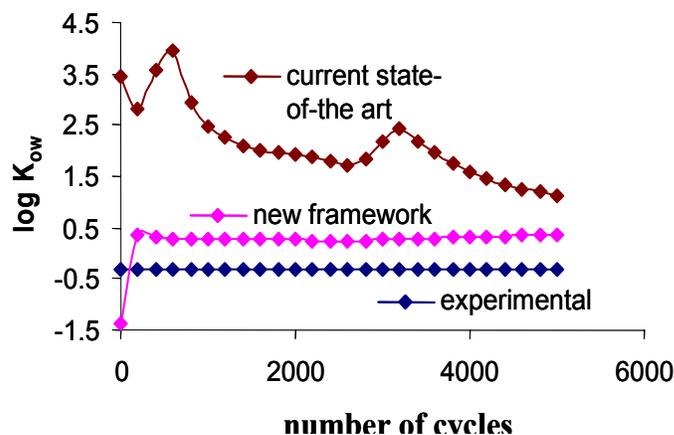


Fig. 15: Octanol-water partition coefficient for propanol predicted by molecular simulations

### Chemical Synthesis

Process design starts with chemical synthesis in a laboratory where a chemical pathway from reactants to products is defined. This involves the search for molecules possessing desired physical, chemical, biological properties. Computer aided chemical synthesis relies on the group contribution methods (77, 78) which assign numerical values to functional groups forming each molecule, through experimental data and theoretical methods. It is possible to calculate a wide range of characteristics for any given chemical by combining these functional groups.

A basic diagram of computer aided molecular design (CAMD) is shown in Fig. 16. The starting point in CAMD is a set of functional groups. All possible combinations of these functional groups are explored to generate molecules that satisfy feasibility constraints. The properties of each group and the interaction parameters between groups can be obtained theoretically, experimentally or by applying statistical regression techniques. Once molecules are generated, the properties of these molecules are inferred from the properties of the functional groups structuring them. If the generated molecule satisfies certain criteria then it is added to the list of candidate molecules. This method can generate a list of candidate molecules for any purpose with reasonable accuracy within a moderate time scale.

There are three main CAMD approaches: generation-and-test, mathematical optimization and

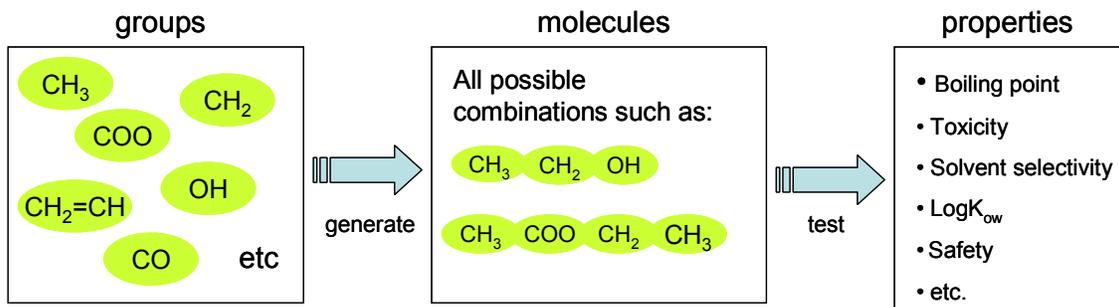


Fig. 16: A basic diagram of computer-aided molecular design

combinatorial optimization approaches (79). All methodologies for CAMD are subject to uncertainties due to experimental errors, imperfect theoretical models/parameters, and inadequate knowledge of the systems. Furthermore, group parameters may not be available and current group contribution models (GCM) cannot estimate all necessary properties.

Uncertainties in CAMD have been addressed in various publications. For example, Maranas (80) studied polymer design with optimal thermophysical and mechanical properties. These properties are estimated based on group contribution methods and there are always discrepancies between the experimental data and the data predicted by group contribution method. In order to model the uncertainties in group contribution parameters, probability distribution functions are utilized, which results in a chance constrained formulation. These chance constraints represent the probability of meeting the target values of properties. The solution to the optimal molecular design problem under uncertainty has at least an  $\alpha$  chance of meeting performance objectives and  $\beta$  chance of maintaining property values within their designated bounds. Since the formulations provided by Maranas (80) involve probability terms, this poses a problem of integrating multivariate probability density distributions. In order to overcome the computational burden, stochastic constraints are transformed into equivalent deterministic ones. This allows reaching an exact solution to the resulting convex MINLP formulation.

Tayal and Diwekar (81) also addressed property prediction uncertainty in polymer design and presented a generalized stochastic framework based on Hammersley Sequence Sampling (HSS) and stochastic annealing to solve this problem. Due to its increased computational efficiency, this framework is applicable to nonlinear or even black box property prediction models, nonlinear objective function and constraints and stable and non-stable distributions for the uncertain variables. It also provides a set of solutions instead of only one solution, which gives flexibility to the designer.

Kim and Diwekar (49, 50, 82) applied the computer-aided molecular design (CAMD) approach to the selection of environmentally benign solvents under uncertainty for extraction. Uncertainties in property prediction models were quantified using available experimental data as shown earlier in Fig. 8. The Hammersley stochastic annealing (HSTA) algorithm was implemented (56) to solve this combinatorial optimization problem. This algorithm makes use of the efficient Hammersley sequence sampling technique for updating discrete combinations, reducing Markov chain length and for determining the number of samples automatically.

The problem of environmentally benign solvent selection was also studied using genetic algorithms (57, 83). Hammersley stochastic genetic algorithm (HSGA) was developed which outperforms the HSTA algorithm by choosing solvents with better targeted properties in less computational time.

### ***Experimental Design, Model Building and Parameter Estimation***

Experimental design is used frequently by researchers for assessing the performance of a new catalyst, determining a reaction mechanism or determining the best operating conditions for a chemical plant. Experimental design also affects the fidelity of the fundamental and semi-empirical models developed and model parameters estimated using experimental data for the process at hand. Reliable models offer competitive advantage to industries for model based process design, operations and control.

In experimental design, it is desired that the experimental region that is sampled generates the maximum amount of information for determining the correct model from a set of candidates and estimating the parameters of this model with the greatest precision.

When building models one uses some prior information such as physical, chemical or biological

laws and propose possible model candidates for the process under consideration. These models may have parameters that have a physical meaning, such as the kinetic constants for a chemical reaction that one wishes to calculate with maximum precision. Usually these models consist of mixed differential and algebraic equation systems.

Model building consists of these three stages (84):

Stage I: Specify one or more models to describe the process and perform preliminary identifiability and distinguishability tests before any data is collected in order to determine whether or not the parameters in the mathematical models can be uniquely identified and model structures can be distinguished from one another.

Stage II: Design experiments for model discrimination to select the best model representative of the process.

Stage III: Design of experiments to improve the precision of the parameters within the best model to arrive at a statistically verified model formulation. The design criterion is to minimize the volume of the confidence region for the parameter estimates.

Once the best model is chosen the parameters of the model are random variables with associated probability distributions. The uncertainty in these parameters affects the predictions made with the models for design, optimization and control. High degree of correlation may also exist within the model parameters. For example, Whiting and coworkers (85) developed a new sampling strategy, called equal probability sampling (EPS), to account for the correlations between thermodynamic model parameters and sample the parameter space more efficiently.

If a complex model is used to represent the physical phenomena, with high number of model parameters, it is time-consuming and expensive to estimate all the parameters with high precision. This task proves especially difficult when there are nonlinearly related model parameters. In these cases, the number of experiments needed to be performed to identify all the model parameters is very costly. Therefore sensitivity analysis may be performed to identify the most significant model parameters and the experimentation can be directed towards determination of these parameters alone, in order to reduce the cost and duration of the parameter estimation and model validation process. Parameter sensitivity analysis is used to quantify the effect of certain model parameters on the model output. Recently, Kontoravdi and coworkers (86) applied Sobol's method for global sensitivity analysis (GSA) to a mammalian cell culture producing monoclonal antibodies and identify the parameters which have a significant impact on the output. Each parameter space is sampled using a Sobol sequence, which yields no overlapping points to thoroughly investigate the entire range of parameter values. This method can be used as a precursor for experimental design to reduce the cost of experimentation. A review of sensitivity analysis techniques for chemical models was presented recently by Saltelli and coworkers (87).

### ***Process Synthesis and Design***

Process synthesis translates chemical synthesis into a chemical process. It encompasses the choice of various unit operations, how they are connected and the optimization of the proposed plant. Process design activities start at this level and a flowsheet of the plant is generated according to these decisions and process simulators are used to predict mass and energy flows for the process. Commonly employed methodologies for selecting optimal process flowsheet configurations can be classified into four groups (74): (1) optimization based approach; (2) hierarchical heuristic approach; (3) thermodynamic phenomena driven approach; (4) evolutionary methods.

Optimization approach to process synthesis involves (a) formulation of a complex flowsheet incorporating all the alternative process configurations, which are called superstructures and (b)



different levels involved in this framework. The innermost level corresponds to models for process simulation. In this level, all possible process alternatives for a particular process are defined. Chemical process simulators such as AspenPlus (95), MultiBatchDS (96) or SuperPro (97) can be used for the innermost modeling. The second level corresponds to the sampling loop where the uncertainties can be specified in terms of probability distributions. Once probability distributions are assigned to the uncertain variables, sampling techniques such as Monte Carlo sampling, Hammersley sampling (HSS) or Latin hypercube sampling (LHS) can be used to perform the sampling operation from the multi-variable uncertain parameter domain as discussed in Section 2. The third level is the continuous optimizer which involves continuous decisions such as process design and operating conditions. In the fourth level, discrete decisions such as chemical and process structural alternatives are managed by mixed integer programs. The outermost loop is the multi-objective programming loop where different objectives such as cost and environmental impacts are considered and trade-off surfaces which are called Pareto optimal solutions are generated. Various applications of this approach such as hybrid fuel cell power plant design under uncertainty (1, 98) and environmentally benign heterogeneous azeotropic distillation system design (68) have been presented. Kheawhom and Hirao (99) also presented a two layer algorithm for environmentally benign process synthesis under uncertainty. In the outer layer, the synthesis problem is represented by a multi-objective optimization problem considering the performances associated with design parameters. In the inner layer, the problem is expressed as a single-objective optimization problem taking in to account the operating performances in the presence of uncertainty. This algorithm was applied to a membrane-based toluene recovery process.

As mentioned earlier, it is essential to identify environmental impacts of a process earlier because the opportunity to overcome environmental problems in later stages of process development diminishes. However, in early design stages, there is high uncertainty in various economical, ecological and technical process parameters. For this purpose, Hoffman and co-workers (100) proposed a new approach to select promising process alternatives in early stages of design. The method is based on approximating flowsheets by polynomial response surfaces with a lower complexity. A multiobjective optimization problem was solved for selecting a production process for hydrocyanic acid with 400 uncertain variables. Latin hypercube sampling technique was performed on the substituted response surface to obtain Pareto optimal solutions.

As stated earlier, process flexibility is an area that received significant attention, as it ensures that processes are operational and safe when exposed to variations in operating conditions. For example for waste reduction in pharmaceutical industries (101, 102) a discrete representation of waste loads assign probabilities to distinct waste scenarios. Since the explicit enumeration of all possible waste scenarios for numerous waste streams would lead to a massive amount of uncertain variables, a randomly selected sample is used to represent the uncertain space based on Monte Carlo sampling. Then a flexibility index is defined which measures the flexibility of a waste treatment policy to changing waste loads and superstructures are searched for recovery and treatment policies. Flexibility issues have also been addressed in process synthesis for heat exchanger network design (103, 104), and synthesis of heat integrated distillation sequences (105). However, these papers use but scenario based approach to represent uncertainties.

### ***Process Operation***

The aim of process operations is to optimally use capital, material, energy and information resources to produce desired chemical products in a reliable and flexible way while complying with environmental and safety regulations. Process operation involves activities ranging from process control, monitoring and diagnosis, process planning, scheduling and supply chain management. The

following subsections describe these various activities. The problems related to production planning, scheduling and supply chain often involve discrete combinatorial decisions and uncertainties. These problems belong to batch processing, which is generally used for the production of high-value added, low volume products such as pharmaceuticals and specialty chemicals. Even though various sources of uncertainties exist in batch processing, most of the literature deal with demand uncertainties. To represent these uncertainties sampling methods are used.

### ***Process Planning, Management and Scheduling***

Most of the problems in management, scheduling and planning include combinatorics (discrete choices and decisions) and uncertainties. Pekny (106) reviewed this area and provided algorithm structures that simultaneously consider combinatorial aspects and data uncertainty for industrial scale problems. The simulation based-optimization approach described in this paper uses a customized mixed integer linear programming solver to optimize process behavior together with a discrete event simulator to investigate the effect of uncertainty on the plans output from the optimizer. This procedure requires a biased sampling scheme to focus on critical events and avoid large number of simulations that are not insightful.

Scheduling problems have been studied widely in chemical engineering literature. Recently, Lin and Floudas (107) presented an overview of scheduling problems in multi-product/multi-purpose batch and continuous processes. The overall profitability of a process and the timely delivery of products highly depend on scheduling. Scheduling problems involve sequencing, assignment of tasks to equipment and maintenance over a planning horizon, and inventory considerations of a process. Scheduling problems could be described conveniently using a resource-task equipment network. These problems are usually encountered in batch processing. In order to determine operating policies based on realistic production plans, uncertain nature of processes must be addressed. In batch processing, uncertainties result from processing time fluctuations, equipment reliability or availability and demand. Two different approaches exist for scheduling problems under uncertainty: (1) reactive scheduling and (2) stochastic scheduling. In reactive scheduling, uncertainties are handled by adjusting the schedule when the uncertain parameters or unexpected events occur. Usually heuristic approaches are used for schedule modifications. Whereas, in stochastic scheduling, the uncertainties are considered at the original scheduling stage and optimal and reliable schedules are found in the presence of uncertainty.

A scenario based approach is usually used which comprises of all the possible future outcomes modeled by discrete probability distributions and the expected value of a performance index such as makespan or profit is optimized with respect to the scheduling decision variables (108). A scenario contains a discrete value for all uncertain variables within a given time interval and its associated probability. The number of scenarios increases exponentially with the number of uncertain variables and this increases the problem size. Bassett and co-workers (109) presented a framework for addressing uncertainties by means of Monte Carlo sampling. Uncertain variable such as processing times and equipment downtimes are sampled from their probability distributions and the reliability of meeting a certain due date is determined. Lee and Malone (110) proposed a probabilistic approach based on the hybridization of Monte Carlo simulation and simulated annealing techniques to obtain a schedule able to handle uncertainties in parameters of batch process scheduling. This approach was also used to develop a flexible planning algorithm (111). Process planning involves the optimal selection of processes from among competing alternatives and timing of capacity expansions in a way that maximizes the net present value of the project over a planning horizon. Liu and Sahinidis (112) developed a two-stage stochastic program for process planning problems under uncertainty using a

combination of Bender's decomposition and Monte Carlo sampling.

### ***Supply Chain Management***

Supply chain management takes the scheduling problems one step further and spans coordination of the management of multiple facilities and shipment of materials through an associated transportation network to customers. Supply chain management spans activities related to storage and movement of raw materials and products from the plant to the point of consumption. Due to the changing market conditions and customer demands, it is critical for businesses to have an efficient and flexible supply chain. Various sources of short-term and long-term uncertainties exist in these systems. Examples of short-term uncertainties are uncertainties in processing parameters such as processing times or yields or availability of equipment. On the other hand, long-term uncertainties include price fluctuations in raw material and final products and seasonal variations in demand, which occur over a longer period of time.

In literature, various sources of uncertainties are addressed in supply chain management. Gupta and Maranas (113) considered demand uncertainty in midterm planning of multisite supply chains. A stochastic programming based approach was described to model the planning process as it reacts to demand realizations unfolding over time. Lababidi and co-workers (114) developed an optimization model to study the supply chain of a petrochemical company under uncertain operating and economic conditions. The objective function is based on optimizing the system resources by minimizing the total production costs and raw material procurement, as well as lost demand, backlog, transportation, and storage penalization. Uncertainties are considered in demands, market prices, raw material costs, and production yields. Multiple scenarios of an uncertain future, each with an associated probability of occurrence were considered. It was found that uncertainties have a profound effect on the planning decisions of the petrochemical supply chain. Jung and co-workers (115) proposed the use of deterministic planning and scheduling models which incorporate safety stock levels as a means of accommodating demand uncertainties in routine operation by a Monte Carlo sampling technique. The problem of determining the safety stock level to use to meet a desired level of customer satisfaction is addressed using a simulation based optimization approach. Wan and co-workers (116) extended the concept of simulation based optimization by introducing a surrogate based model together with domain reduction and incremental sampling to extract structure information from noisy simulation results and to optimize supply chain decisions. The idea behind a surrogate-based model is to fit a single surface for the whole decision space, and use this surface to perform optimization instead of the simulation model. This model is constructed using Latin hypercube sampling (LHS), domain reduction techniques to concentrate on the exploration of good regions, and support vector machines to extract structure information from noisy data.

Guillen and co-workers (117) presented a stochastic multi-objective optimization approach to obtain a trade-off between customer satisfaction and expected profit to be achieved in the short term operation of chemical industry supply chains. A two-stage stochastic formulation is used, which considers the uncertainty associated with reactions to future demand and a set of Pareto optimal solutions are generated. This approach is aimed to provide decision support in making optimal offer proposals during negotiation process between customers and suppliers. The uncertainty associated with product demands and prices is represented by a set of scenarios with a given probability of occurrence and these scenarios are generated by performing Monte Carlo sampling.

Hung and co-workers (118) presented a new modeling approach based on an object-oriented architecture to handle supply chain configurations, operational decisions and policies, through the use of a generic supply chain node. The model provides a fully dynamic simulation of the supply-chain

and the effect of various uncertainties are evaluated through Monte Carlo simulation and other more efficient, sampling techniques based on quasi-Monte Carlo methods. The uncertain variables in supply chains are sampled from their respective probability distributions and the expected value of a performance indicator such as customer service level or average inventory is calculated

### ***Reliability***

Due to increased competition worldwide, chemical plants need to operate with high process reliability to increase operational effectiveness and profits. System reliability and availability methods can be classified as measurement based and model based methods (119). Measurement based methods are expensive as they require building a real system or its prototype and taking measurements and then analyzing the data statistically. In the context of process systems, at the design stage where the system or its prototype is not yet been built, the use of measurement technique is nor feasible. While at the operational stage, it can prove to be very expensive to inject faults into a real system to measure data. Model based methods are much easier to use and are particularly useful at the design stage to screen lots of design alternatives without building the actual system. However, model-based methods are subject to model uncertainties, which propagate into RAM (reliability, availability and maintainability) performance (120)

It has become important in recent years to address reliability issues at the conceptual design stage. It is also critical to increase the availability of the plant to save on lost production costs. The problem of including uncertainties in equipment availability at the design stage was addressed by Pistikopoulos and coworkers (121-124) Obtaining an optimal production schedule in the presence of equipment failure uncertainty for multiproduct/multipurpose batch plants is important for profitability and timely production. Sanmarti et al. (125) addressed this problem and introduced a schedule reliability index to identify robust schedules. This reliability index represents the represents the discrete probability that a corresponding unit will be available to perform the next scheduled task based on the failure history and maintenance operations carried out on the unit. Production and maintenance schedules were determined simultaneously with this methodology. However, only scenario based approach is used for the probabilistic evaluation of reliability and availability in all these papers.

Genetic algorithms were applied to preventive maintenance optimization problems by Tan and Kramer (126). Their framework for preventive maintenance optimization combines Monte Carlo simulation with a genetic algorithm. This framework is suitable for handling uncertainties and non-deterministic objective functions.

### ***Risk Analysis and Research Management***

Risk and policy analysis involves uncertainty quantification and characterization using probability distributions and sampling. Since the results of the probabilistic analysis depend on the number of samples chosen, the choice of an efficient sampling technique becomes crucial. It is desirable to use a sampling technique that can predict the output probabilistic measure accurately with the minimum number of samples. Wang and co-workers (25) presented new sampling techniques based on the combination of Hammersley sequence sampling (HSS) and Latin hypercube sampling (LHS) for the evaluation of health risk associated with exposure to hazardous materials. This sampling technique inherits the advantages of both HSS and LHS for superior efficiency.

Sampling techniques are also used for financial risk assessment in chemical process industries. For example, Bonfill and co-workers (127) presented a stochastic optimization approach to manage risk in short-term scheduling of multi-product batch plants with demand uncertainties. A two-stage

stochastic optimization model accounting for the maximization of the expected profit was used and this model was also extended to incorporate the availability of option contracts. To represent the demand uncertainty, independent scenarios were simulated by Monte Carlo sampling from the given probability distributions. A similar approach was applied also for processing time uncertainties (128). Guillen et al. (129) developed a new strategy for integrating pricing decisions with the scheduling of batch plants to manage financial risk associated with demand uncertainty. The relationship between prices and demand have been modeled and forecasted and integrated into the scheduling model to determine simultaneously the prices and optimal schedule to maximize the profit. A sample average approximation (SAA) method was used to approximate the expected profit in the objective function.

Research management in general is related to research prioritization and reduction of uncertainties. A “value of research” methodology was proposed by Johnson and Diwekar (92) and Johnson and co-workers (130) for research management problems. This methodology tries to determine when imperfect information is acceptable, and where should the scarce resources be allocated to leverage the impact of these research efforts on the whole of its strategy. While reducing uncertainty is profitable, the time required to achieve a reduction tempers the benefit. This approach was applied recently to hybrid fuel cell power plants (131).

### ***Robust Control***

Control systems are used to keep the product specifications on target, to minimize deviations from the nominal process conditions and maintaining the safe operation of the plant. Control system design involves the selection of input and output variables, the process model, appropriate type of controller and adjustment of the controller tuning parameters. Model uncertainty and external disturbances are important concerns in designing control systems.

One of the most important criteria for designing a control system is to achieve robustness to these model uncertainties and disturbances. A probabilistic approach has been used by Schaper and co-workers (132) to achieve robust process control. This probabilistic approach characterizes the model uncertainties by probability distributions and a statistical measure of disturbance rejection for the controller is incorporated into a robust control framework. The performance of the controller is then characterized by a probability measure for all situations between nominal and worst case conditions. Ratto and Paladino (133) considered uncertainties in process and kinetic parameters for non-ideal controlled CSTRs and described a procedure to perform stability, sensitivity and bifurcation analysis by a Monte Carlo method. This procedure is used to identify most probable stability regions and to design a robust control system. Li and co-workers (134) proposed a model predictive control strategy under chance constraints for robustness. Both the model and disturbance uncertainties were considered and assumed to be correlated multivariate stochastic variables. A stochastic program under joint probabilistic constraints was formulated and using the Hammersley Sequence Sampling (HSS) technique, this problem was relaxed to a nonlinear programming problem.

In order to achieve robustness, parameter design methodology is also a widely used method. It is termed as an off-line quality control method for designing products and manufacturing processes that are robust in the face of uncontrollable variations popularized by Taguchi (135). The variables affecting a product’s performance are classified into two groups: (1) design parameters whose nominal settings can be specified; (2) noise parameters that represent uncontrollable variations over a product’s lifetime and across different units. In order to relate the noisy input parameters to the process output, two different approaches could be used: (1) physical experiments could be conducted by varying the input parameters over the noise space to generate a response surface or (2) computational models could be developed. Monte Carlo methods are used for propagating the effects of input variability through a

model and output variability is studied. A sample of input vectors is generated which is representative of the uncertainty distribution and outputs are evaluated at each of these samples.

The importance of sampling efficiency for generating these samples from the multivariate space for the formulation of a stochastic optimization problem for parameter design was emphasized in an earlier study by Diwekar and Rubin (90) for off-line quality control of a continuous stirred tank reactor. Latin Hypercube Sampling (LHS) was used instead of Monte Carlo technique to reduce the required number of samples. Later, Kalagnanam and Diwekar (13) applied the Hammersley Sequence Sampling (HSS) technique to this problem for further efficiency improvements exploiting the k-dimensional uniformity properties of this technique. Another study related to robust batch distillation column design using the HSS technique also illustrated the usefulness of this approach. Sahin and Diwekar (136) demonstrated efficiency of a new algorithm called Better Optimization of Nonlinear Uncertain Systems (BONUS) for the same problem. Terwiesch and Agarwal (137) presented an optimization procedure to achieve robustness in batch reactor optimal control under parametric uncertainties. Probability distributions were used for the uncertain process parameters and the expectation of cost function for the entire parameter space was optimized.

### ***Optimal Control***

A challenging control problem in which has received considerable attention in literature is optimal control where an optimal trajectory (future of action) for a control variable is computed by dynamic optimization, so as to maximize/minimize a performance index such as cost, product yield or time. To compute optimal control trajectories in the face of time-dependent uncertainties, a stochastic maximum principle formulation was developed based on real options theory and using a class of stochastic processes called *Ito processes* (51, 138). Using this methodology, thermodynamic model uncertainties in batch distillation were characterized and stochastic optimal control profiles were computed (47, 51, 139). Sampling techniques and stochastic modeling approaches were used to characterize, quantify and propagate uncertainties. This technique improved the process performance objectives significantly in various case studies.

In recent years considering process design and control simultaneously has become an important. Sakizlis and co-workers (140) have provided a review of recent advances towards the integration of process design, process control and process operability where time-varying disturbances and uncertainties are considered. More recently, Pajula and Ritala (141) discussed the effects of measurement uncertainty on process performance and how it should be accounted for in the design of the control structure. For this purpose, dynamic scenarios were used and they were each assigned a probability of occurrence using the knowledge gained from earlier experiences. This method was applied to a papermaking process where the effect of measurement uncertainty of fiber and filler consistency (concentration) on controller performance was studied. For batch separations and solvent recycling in pharmaceutical industry, Ulas and Diwekar (142) presented a framework that couples product design, process design and optimal control in the face of time-dependent uncertainties.

### **Conclusion and Future Trends**

Sampling is an important element of uncertainty analysis, stochastic modeling and optimization algorithms, used for chemical process design, operation and control. Sampling techniques are employed to sample probabilistic space of uncertain variables commonly encountered in these applications. Apart from uncertainty analysis, sampling also plays an important role in improving efficiency of discrete, stochastic and multi-objective optimization algorithms. Sampling based Monte Carlo methods are also an essential part of computational chemistry.

Monte Carlo sampling is the most commonly used sampling technique based on pseudorandom numbers. This sampling technique has probabilistic error bounds and it requires large sample sizes to estimate the mean and standard deviation for an uncertain variable. Therefore, variance reduction techniques have been employed in order to increase efficiency. Importance sampling, Latin hypercube sampling (LHS) and Hammersley sequence sampling (HSS) are examples of variance reduction techniques. Importance sampling is executed on the fact that some of the input random variables have more impact on the parameter being estimated than others and, these values are sampled more frequently. On the other hand, Latin hypercube sampling is a stratified sampling technique to reduce the sample size. Hammersley Sequence Sampling (HSS) and its variants are efficient sampling techniques based on quasi-random numbers showing k-dimensional uniformity properties.

Due to increased environmental consciousness, traditional design methods should include objectives such as environmental and health impacts, risk, reliability and safety, as well as controllability and profitability into earlier stages of process design. Sampling techniques have various applications during the life cycle of the plant such chemical synthesis, process synthesis, process operations such as management and planning, supply chains, scheduling, control and maintenance optimization for better reliability. Environmental and financial risk management are other applications where sampling is a crucial step.

Future trends in sustainable process design require researchers to study the connections between industries and ecosystems, which are the complex networks of humans, plants, animals and the environment (2). The effects of hazardous chemicals and the activities of the chemical plants with the ecosystem need to be modeled for enhanced decision making. These multifaceted models have many steady state and dynamic uncertainties and efficient sampling techniques will play an important role in analysis.

Furthermore in the future, the applicability of efficient sampling techniques such as Hammersley sequence sampling (HSS), needs to be examined for higher dimensional problems. There is a loss of uniformity observed in quasi-random sequences for higher dimensions. Leaping techniques (24, 26) and combinations of quasi-random sequences are used to improve Hammersley sequence sampling (HSS) technique to make it applicable to higher dimensional problems and more work is needed in this area.

## References

1. U.M. Diwekar, *Environmental Science and Technology* **37** (23): 5432-5444. (2003)
2. U. Diwekar, *Resources, Conservation and Recycling* **44** (3):215-235 (2005)
3. N. Metropolis and S. Ulam, *Journal of American Statistical Association*. **44** (247), 335-341 (1949)
4. D.H. Lehmer, *Proceedings of the 2<sup>nd</sup> Symposium on Large Scale Digital Calculating Machinery*, 141-146 (1949)
5. B. Wichmann and I. Hill, *Applied Statistics* **31**: 188-190 (1982).
6. B.A.P. James, *The Journal of the Operational Research Society* **36**(6), 525-530 (1985).
7. B.L. Nelson, *Computers and Operations Research*. **14**: 218-225 (1987)
8. B.L. Nelson, *Operations Research* **38** (6): 974-992. (1990)
9. J.R. Wilson, *Amer. J. Math. Management Sci.* **3**: 121-144.(1983)
10. M.D. McKay., R.J. Beckman, W.J. Conover, *Technometrics*, **21**(2): 239-245.(1979)
11. R.L. Iman and W.J. Conover, *Communications in Statistics* **A17**:1749-1842.(1982)
12. E. Saliby, *Journal of Operations Research Society* **41**(12), 1133-1142 (1990).
13. J.R. Kalagnanam and U.M. Diwekar, *Technometrics* **39** (3): 308-319 (1997)
14. U.M. Diwekar and J.R. Kalagnanam, *AIChE Journal* **43**(2): 440-447.(1997)

15. D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, San Diego: Academic Press (1996).
16. D.E. Knuth, *The Art of Computer Programming Volume 1: Fundamental Algorithms* Addison-Wesley Publishing Co (1973).
17. D.A. Straub and I.E. Grossmann, *Computers & Chemical Engineering* **14** (9): 967-985 (1990)
18. F.P. Bernardo, E.N. Pistikopoulos and P.M. Saraiva, *Computers & Chemical Engineering* **25** (1): 27-40 (2001)
19. M. Wendt, P. Li, G. Wozny, *Industrial & Engineering Chemistry Research* **41**(15): 3621-3629 (2002)
20. Vasquez V.R. and Whiting W.B, *Computers & Chemical Engineering* **23** (11-12): 1825-1838 (2000)
21. H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*. SIAM Publications, Philadelphia, PA. (1992)
22. J. H. Halton, *Numerische Mathematik* **2**: 84-90 (1960)
23. J.M. Hammersley, *Annals of the New-York Academy of Science* **86** : 844-874. (1960)
24. K. Subramanyan and U.M. Diwekar, *User Manual for Fortran-Based Stochastic Sampling Code*. (2006)
25. R.Y. Wang, U.M. Diwekar, C.E.G. Padro, *Environmental Progress* **23** (2): 141-157 (2004)
26. L. Kocis and W. Whiten, *ACM Transactions on Mathematical Software* **23** (2): 266-294. (1997)
27. P. Hellekalek, Proceedings of the 12<sup>th</sup> workshop on Parallel and distributed simulation pp: 82-89, IEEE (1998)
28. M. Mascagni, "Some Methods of Parallel Pseudorandom Number Generation," in *Algorithms for Parallel Processing*, R. Schreiber, M. Heath and A. Ranade editors, Springer Verlag: New York, Berlin, pp. 277-288 (1997)
29. M. Mascagni, "Deterministic Monte Carlo Methods and Parallelism," *Sourcebook on Parallel Computing*, J. Dongarra, I. Foster, F. Fox, W. Gropp, K. Kennedy, L. Torcson, and A. White, editors, Morgan Kaufman Publishers, San Francisco, pp. 249-258 (2003)
30. W.C. Schmid and A. Uhl, *Mathematics and Computers in Simulation* **55** (1-3): 249-257 (2001)
31. T. Bayes, *Phil. Trans. R. Soc. Lond.* **53**: 370-418 (1763)
32. A. Meel, L.M. O'Neill, W.D. Seider, U. Oktem and N. Keren, *AIChE Spring National Meeting*, Paper 106d. (2006)
33. N. Mehranbod, M. Soroush and C. Panjapornpon, *Journal of Process Control* **15**(3):321-339. (2005)
34. Y. Bard, *Nonlinear Parameter Estimation*, Academic Press, New York, pp. 72-74 (1974).
35. S. Geman, and D. Geman, *IEEE Trans. Pattern Anal. Mach. Intell.* **6**: 721-741 (1984)
36. N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, *J. Chem. Phys.* **21**: 1087-1091 (1953)
37. W.K. Hastings, *Biometrika* **57**: 97-109 (1970)
38. S.P. Brooks, *The Statistician* **47** (Part 1): 69-100 (1998)
39. W.R. Gilks, S. Richardson, D.J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Chapman and Hall, Great Britain (1996)
40. W.G. Madow and L.H. Madow, *Annals of Mathematical Statistics* **15**(1):1-24. (1944)
41. S.K. Thompson, *Journal of the American Statistical Association* **85** (412): 1050-1059 (1990)
42. C. Laurent and J.F. Etard., *Revue D'Epidemiologie et de Sante Publique* **53** (1): 43-50 (2005)
43. P. Turk and J.J. Borkowski. *Environmental and Ecological Statistics* **12** (1): 55-94 (2005)
44. T.M.F. Smith, *Journal of the Royal Statistical Society. Series A (General)*, **146** (4): 394-403 (1983)
45. A.S. Whittemore, *J. Halpern Statistics in Medicine* **16** (1-3): 153-167. (1997)
46. A. Chaudhuri, A.K Adhikary., S. Dihidar, *Metrika* **52** (2): 115-131. (2000).
47. S. Ulas, and U.M. Diwekar, *Computers & Chemical Engineering* **28** (11):2245-2258. (2004)
48. U.M. Diwekar, and E.S. Rubin, *Computers & Chemical Engineering* **15** (2):105-114. (1991)
49. K.-J. Kim and U.M. Diwekar, *Ind. Eng. Chem. Res.* **41** (5):1276-1284. (2002)
50. K.-J. Kim and U.M. Diwekar *Ind. Eng. Chem. Res.* **41** (5): 1285-1296. (2002)
51. V. Rico-Ramirez, U.M. Diwekar, B. Morel *Computers & Chemical Engineering* **27** (12): 1867-1882 (2003)

52. U.M. Diwekar, *Introduction to Applied Optimization*. Boston, MA: Kluwer Academic Publishers (2003)
53. A.K. Dixit, and R.S. Pindyck, *Investment under Uncertainty*. Princeton, NJ: Princeton University Press (1994)
54. P. J. M. VanLaarhoven and E. H. Aarts *Simulated Annealing Theory and Applications*, D. Reidel Publishing Co: Holland. (1987)
55. L. A. Painton, and U.M. Diwekar, *European Journal of Operational Research* **83**: 489-502. (1995)
56. K.-J. Kim and U.M. Diwekar *IIE Transactions* **34** (9): 761-777. (2002).
57. U.M. Diwekar, and W. Xu, *Ind. Eng. Chem. Res.* **44** (18):7132-7137. (2005)
58. J.R. Birge and F. Louveaux, *Introduction to Stochastic Programming*. Springer Series in Operations Research, Springer-Verlag New York, Inc. (1997)
59. J.R. Birge and F. Louveaux *European Journal of Operations Research* **34**:384-392. (1988)
60. G.B. Dantzig and G. Infanger, *In: Brezinski C. and U. Kulisch (Eds.) Computational and Applied Mathematics I-Algorithms and Theory*. Proceedings of the 13th IMCAS World Congress, Dublin, Ireland, 111-120. (1991)
61. J.L. Higle and S. Sen *Mathematics of Operations Research* **16**(3): 650-669.(1991)
62. J. Wei, and M.J. Realff, *Computers & Chemical Engineering* **28** (3):333-346. (2004)
63. P.D. Chaudhuri, and U.M. Diwekar, *AIChE Journal* **42**, 742-752. (1996)
64. P. Chaudhuri, and U. Diwekar, *AIChE Journal* **45** (8):1671-1687. (1999)
65. S. Sambarey and U.M. Diwekar *Grace Hopper Celebration of Women in Computing*, San Diego, California (2006)
66. C. Hwang, S. Paidy, and K. Yoon *Computers & Operations Research* **7**: 5-31 (1980)
67. Y. Fu, and U.M. Diwekar, *Annals of Operations Research* **132**, 109-134. (2004)
68. W. Xu and U.M. Diwekar, *Ind. Eng. Chem. Res.* **44** (11):4061-4067 (2005).
69. Leach A. *Molecular Modeling Principles and Applications*, Prentice Hall, Pearson Education Limited, Great Britain. (2001)
70. Siepmann, J. I., and D. Frenkel. *Molecular Physics* **75**(1): 59-70. (1992)
71. T.J.H. Vlugt, M.G. Martin, B. Smit, J.I. Siepmann, and R. Krishna, *Molecular Physics* **94** (4):727-733. (1998).
72. M.G. Martin, and J. I. Siepmann, *Journal of Physical Chemistry B*, **103**(21):4508-4517 (1999)
73. Z. Chen, and F. A. Escobedo, *Journal of Chemical Physics* **113** (24):11382-11392 (2000)
74. Y. Okamoto, , *Journal of Molecular Graphics & Modeling*, **22** (5):425-439 (2004)
75. A. Mitsutake, Y. Sugita, Y. Okamoto, *Biopolymers* **60** (2): 96-123. (2001)
76. S. Ulas and U.M. Diwekar, *Molecular Simulation* (accepted, invited paper for the special issue of *Frontiers in Molecular Simulations*) (2006b)
77. K.G. Joback and R.C Reid, *Chemical Engineering Communications* **57** (1-6): 233-243 (1987)
78. H.K. Hansen, P. Rasmussen, A. Fredenslund, M. Schiller, J. Gmehling, *Industrial & Engineering Chemistry Research* **30** (10): 2352-2355 (1991).
79. P.M. Harper, G. Rafiqul, P. Kolar and T. Ishikawa, *Fluid Phase Equilibria* **158-160**: 337-347 (1999).
80. C.D. Maranas, *AIChE J.* **43** (5): 1250-1264. (1997)
81. M.C. Tayal and U.M. Diwekar *AIChE J.* **47**(3):609-628 (2001)
82. K.-J. Kim, and U.M. Diwekar, *Ind. Eng. Chem. Res.* **41** (18):4479-4488. (2002)
83. W. Xu and U.M. Diwekar, *Ind. Eng. Chem. Res.* **44**(18); 7138-7146. (2005)
84. S.P. Asprey and S. Macchietto, *Computers and Chemical Engineering* **24**: (2-7): 1261-1267.
85. W.B. Whiting, V.R. Vasquez and M.M. Meerschaert, *Fluid Phase Equilibria* **158-160**: 627-641 (1999)
86. C. Kontoravdi, S.P. Asprey, E.N. Pistikopoulos, A. Mantalaris, *Biotechnology Progress* **21**(4): 1128-1135 (2005)
87. A. Saltelli, M. Ratto, F. Campolongo, *Chemical Reviews* **105** (7): 2811-2828 (2005)
88. U.M. Diwekar, E.S. Rubin, and H.C. Frey, *Energy Conversion and Management* **38**, 1725-1735. (1997)
89. U.M. Diwekar, *AIChE Symposium Series on Pollution Prevention through Process and Product Modifications* **90**(303):168-79 (1995)

90. U.M. Diwekar and E.S. Rubin, *Ind. Eng. Chem. Res.* **33**(2): 292-298 (1994)
91. V. Narayan,, U.M. Diwekar, and M. Hoza, *Ind. Eng. Chem. Res.* **35**(10); 3519-3527 (1996)
92. T.L. Johnson, and U.M. Diwekar, In *Foundations of Computer-Aided Design*; (Eds. Malone, M. F., Tainham, J. A., Carnahan, B.), AIChE Symposium Series **96**: 454-461 (2000).
93. M.M. Dantus and K.A. High *Computers and Chemical Engineering* **23** (10): 1493-1508 (1999)
94. J. Acevedo, and E.N. Pistikopoulos, *Computers & Chemical Engineering* **22**, 647-671. (1998)
95. Aspen Tech, AspenTech. Aspen Plus 12.1 User Guide. Cambridge, MA (2003)
96. U.M. Diwekar, *Comput. Appl. Chem. Eng. Educ.* **4** (4): 275-84. (1996)
97. T. Shanklin, K. Roper, P.K. Yegneswaran, M. Marten, *Biotechnol. Bioeng.* **72** (4):483-489 (2001)
98. K. Subramanyan, U. Diwekar, A. Goyal, *Journal of Power Sources* **132** (1-2): 99-112 (2004)
99. S. Kheawhom, and M. Hirao, *Computers & Chemical Engineering* **28** (9):1715-1723. (2004)
100. V.H. Hoffmann, G.J. McRae, and K. Hungerbuhler, *Ind. Eng. Chem. Res.* **43** (15):4337-4349 (2004)
101. A.A Linninger, A. Chakraborty, and R.D. Colberg, *Computers & Chemical Engineering* **24** (2-7):1043-1048. (2000)
102. A.A. Linninger, and A.Chakraborty, *Computers & Chemical Engineering* **25** (4-6):675-681 (2001)
103. L. Tantimuratha, and A.C. Kokossis, *Annals of Operations Research* **132**, 277-300 (2004)
104. Z.N. Pintaric, and Z. Kravanja, , *Computers & Chemical Engineering* **28** (6-7):1105-1119. (2004).
105. G.E. Paules, and C.A. Floudas, *Computers & Chemical Engineering* **16**, 189-210. (1992)
106. J.F. Pekny *Computers & Chemical Engineering* **26** (2): 239-267 (2002)
107. X. Lin, S.L. Janak, and C.A. Floudas, FOCAPO 2003 Special issue. *Computers & Chemical Engineering* **28** (6-7):1069-1085. (2004)
108. M.G. Ierapetritou and E.N. Pistikopoulos *Ind. Eng. Chem. Res.* **35** (3): 772-787 (1996)
109. M.H. Bassett, J.F. Pekny, and G.V. Reklaitis, *Computers & Chemical Engineering* **21** (Supplement 1):S1203-S1208 (1997)
110. Y.G. Lee and M.F. Malone *International Journal of Production Research* **39**(4): 603-623. (2001)
111. Y.G. Lee, and M.F. Malone, , *Ind. Eng. Chem. Res.* **40** (6): 1507 – 1515 (2001)
112. M.L. Liu, and Sahinidis, *Ind. Eng. Chem. Res.* **35** (11) 4154 – 4165 (1996)
113. A. Gupta, and C.D. Maranas, , *Computers & Chemical Engineering* **27** (8-9):1219-1227. (2003)
114. H.M.S. Lababidi, M.A. Ahmed, I.M Alatiqi., A.F. Al-Enzi, *Ind. Eng. Chem. Res.* **43**(1): 63-73 (2004)
115. J.Y. Jung, G. Blau, J.F. Pekny, G.V. Reklaitis, and D. Eversdyk, , *Computers & Chemical Engineering* **28** (10):2087-2106. (2004)
116. X. Wan, J.F. Pekny, and G.V. Reklaitis,, *Computers & Chemical Engineering* **29** (6):1317-1328 (2005)
117. G. Guillen, F.D. Mele, M.J. Bagajewicz, A. Espuna, and L. Puigjaner, , *Chemical Engineering Science* **60** (6):1535-1553. (2005)
118. W.Y. Hung, N.J. Samsatli and N. Shah *European Journal of Operational Research* **169** (3): 1064-1076 (2006)
119. A. Sathaye, S. Ramani, and K.S. Trivedi, *Proceedings of International Workshop on Fault-Tolerant Control and Computing (FTCC-1)* (2000)
120. H. Goel, *Integrating Reliability, Availability and Maintainability (RAM) in Conceptual Process Design, An Optimization Approach*. Delft University Press, Delft, Netherlands (2004)
121. E.N. Pistikopoulos, T.V. Thomaidis, A. Melin, and M.G. Ierapetritou, *Computers & Chemical Engineering* **20**, S1209-S1214 (1996)
122. E.N. Pistikopoulos, C.G. Vassiliadis, and L.G. Papageorgiou, *Computers & Chemical Engineering* **24**, 203-208 (2000)
123. E.N. Pistikopoulos, C.G. Vassiliadis, J. Arvela, and L.G. Papageorgiou, *Industrial & Engineering Chemistry Research* **40**, 3195-3207 (2001)
124. C.G. Vassiliadis, and E.N. Pistikopoulos, *Computers & Chemical Engineering* **23**, S555-S558 (1999)
125. E. Sanmarti, A. Espuna, and L. Puigjaner, *Computers & Chemical Engineering* **21**, 1157-1168 (1997)
126. J.S. Tan, and M.A. Kramer, *Computers & Chemical Engineering* **21**, 1451-1469 (1997)

127. A. Bonfill, M. Bagajewicz, A. Espuna, L. Puigjaner, *Ind. Eng. Chem. Res.* **43** (3): 741-750 (2004)
128. A. Bonfill, A. Espuna, L. Puigjaner, *Ind. Eng. Chem. Res.* **44** (5): 1524-1534 (2005)
129. G. Guillen, M. Bagajewicz, S.E. Sequeira, A. Espuna, L. Puigjaner, *Ind. Eng. Chem. Res.* **44** (3):557-575 (2005)
130. T. L. Johnson, and U. M. Diwekar, *J. Multi-Criteria Decision Analysis* **10**, 87-99. (2001)
131. K. Subramanyan and U.M. Diwekar *Industrial & Engineering Chemistry Research* **45** (2): 681-695 (2006)
132. C.D. Schaper, D.E. Seborg, D.A. Mellichamp *Industrial & Engineering Chemistry Research* **31** (7): 1694-1704 (1992)
133. M. Ratto, and O. Paladino, *Chemical Engineering Journal* **79** (1):13-21. (2000)
134. P. Li, M. Wendt, and G. Wozny, *Computers & Chemical Engineering* **24** (2-7):829-834. (2000)
135. G. Taguchi and Y. Wu *Introduction to Off-line Quality Control*, Nagoya, Japan: Central Japan Quality Control Association. (1980)
136. K. Sahin and U. Diwekar, *Annals of Operations Research*, 132, 47 (2004)
137. P. Terwiesch, M. Agarwal *Chemical Engineering Communications* **131**: 33-52 (1995)
138. V. Rico-Ramirez and U.M. Diwekar *Computers & Chemical Engineering* **28** (12): 2845-2849 (2004)
139. S. Ulas, U.M. Diwekar, and M.A. Stadtherr, *Computers & Chemical Engineering* **29** (8):1805-1814 (2005)
140. V. Sakizlis, J.D. Perkins, E.N. Pistikopoulos *Computers & Chemical Engineering* **28** (10): 2069-2086 (2004)
141. E. Pajula, and R. Ritala, *Chemical Engineering and Processing* **45** (4):312-322. (2006)
142. S. Ulas and U.M. Diwekar *Chemical Engineering Science* **61** (6): 2001-2009 (2006)