Korean J. Chem. Eng., **19**(4), 535-544 (2002)

Improved Evolutionary Operation Based on D-optimal Design and Response Surface Method

Young-Hwan Chu, In-Su Han and Chonghun Han†

Department of Chemical Engineering, Pohang University of Science and Technology, Pohang, Kyungbuk 790-784, Korea
(*Received 3 December 2001 • accepted 1 April 2002*)

Abstract−This paper presents improved evolutionary operation based on D-optimal design and response surface method. D-optimal design and response surface method allow us to overcome the disadvantages of conventional evolutionary operation. Although evolutionary operation has been an effective alternative when fundamental models are hard to build because of the lack of the necessary information, the disadvantages in the number of experiments, experimental design and analysis and detection of the optimal point have prevented EVOP from being frequently applied to real processes. To compare the performance of the proposed method and the conventional EVOP, both of them were applied to a pulp digester process. As a result, the comparable response variable value has been clearly obtained with the proposed method while conducting much fewer numbers of experiments than the conventional evolutionary operation. In addition, the proposed method flexibly handles the constraints in the experimental design and gives more reliable experiment results than the conventional evolutionary operation. By virtue of these benefits, the proposed method can be utilized effectively for a process where prior knowledge for the process is not available.

Key words: Optimization, EVOP, D-optimal Design, Response Surface Method, Pulp Digester Process

INTRODUCTION

It has been indispensable for industry to cut down operating costs and to enhance the quality of products to survive in extreme competition. To achieve these goals, we should find the optimal operating conditions at which to operate a plant. Although industrial operators have improved operating conditions based on their previous experiences, their operating methods might be neither systematic nor economical. Thus, it usually takes a great deal of time to reach the optimal operating condition and it may not be the global optimum, even if they have found out an improved point. To solve this problem, various optimization methods based on mathematical programming have been proposed [Lee and Lim, 1999; Choi et al., 2000; Janson, 2001; Choi and Manousiouthakis, 2002]. The methods necessarily include first principle models for a process as equality constraints in an optimization. Therefore, only if the models can describe the given system to an acceptable degree we can use these methods to effectively locate the optimum point. On the other hand, the necessity for the accurate first principle models restricts usefulness of the methods because it is very difficult and time-consuming to build the first principle models for an unknown process whose static and dynamic behaviors cannot be exactly modeled. For this reason, the mathematical programming based methods have been usually applied to relatively simple and well-known processes such as utility plants [Lee et al., 1998a, b; Yi and Han, 2001].

Evolutionary operation (EVOP) proposed by Box [1957] can be applied for this case since it allows us to find the optimal operating condition without using the first principle models. There have been many cases where EVOP was applied for the optimization of chemical processes [Hunter and Kittrell, 1966]. Although the application cases of EVOP have rapidly decreased since the 1970s because of advances in modeling techniques, EVOP has many useful properties as an optimization method. In particular, it is advantageous to the optimization of complex processes since EVOP does not require accurate first principle models for a process.

This is proved by the fact that EVOP is again being used for the optimization of bioprocesses for which the inner phenomena are not clearly understood. Banerjee and Bhattacharyya [1993] applied EVOP to a bioprocess, where information on the process was insufficient, to maximize enzyme activity using three inducers. Tunga et al. [1999] also used EVOP to maximize the production of protease by optimizing the concentrations of vitamin, metal ion and plant hormone. Saad [1994] showed that EVOP could be used for the optimization of the porcelain enamel manufacturing process. All of them mentioned that EVOP could be applied to unknown processes as a superb optimization technique. However, several disadvantages of EVOP, such as a large number of experiments, ambiguousness of the analysis result on the direction of process improvement, and excessive reduction of experiment region in the presence of constraints, should be overcome for its more frequent applications to real problems.

Consequently, in this paper, we propose an improved EVOP that overcomes the limitations of a conventional EVOP. In the proposed method, D-optimal design and response surface method (RSM) are used to solve the problems. D-optimal design, which is known as one of the most efficient experimental design methods, has been greatly developed by Mitchell [1974], Johnson and Nachtsheim [1983], and DuMouchel and Jones [1994]. Mitchell [1974] proposed DETMAX algorithm which is known as the origin of the modern D-optimal design algorithm. Johnson and Nachtsheim [1983] suggested some guidelines to construct exact D-optimal designs on convex design spaces. They recommended Galil and Kiefer's

†To whom correspondence should be addressed.
E-mail: chan@postech.ac.kr

method [Galil and Kiefer, 1980] and Powells method [Powell, 1964] to determine a good starting design and to efficiently find a D-optimal experiment subset from the viewpoint of optimization, respectively. D-optimal design was further advanced by DuMouchel and Jones [1994]. They modified the algorithm of D-optimal design by combining Bayesian paradigm with the notion of primary and potential terms to make it more resistant to the biases caused by an incorrect model. Response surface method (RSM) for building an empirical model using experiment data was first addressed by Box and Wilson [1951], and then has been utilized in many fields including chemical engineering, bioengineering, pulp and paper industry, and pharmaceutical industry since it is a well-established method for investigating the causal relationship between inputs and outputs for a system [Park et al., 1996].

In this study, first, we explain how the problems of the conventional EVOP are solved by D-optimal design and response surface method. Then, we compare the performance of the proposed method with that of the conventional EVOP by applying both methods to the pulp digester benchmark model [Kayihan, 1997].

THEORETICAL BACKGROUND

1. EVOP and Its Improvement Based on D-optimal Design and RSM

EVOP is a method for process improvement proposed by Box in 1957 [Box, 1957]. The basic philosophy of EVOP is that industrial processes should be run so as to generate not only products but also the information on how the product can be improved. By applying EVOP to a process, operators explore an unexperienced operating region on the basis of results of 2^n factorial design of experiments. And then, they move an operating condition to a better point by statistically investigating the effects of input factors. EVOP is implemented in an actual process itself as an operation mode. That is, it is virtually a permanent method of running a plant. Therefore, it does not require special facilities and concessions.

Basically, a conventional EVOP can be carried out by following several steps as shown in Fig. 1. Of the steps, the 2^n factorial design [Box and Draper, 1987] as an experimental design method and statistical test procedure diminish the usefulness of EVOP and prevent it from being frequently applied to real plants as an optimization method in spite of its many advantages. First, the 2^n factorial design used in the conventional EVOP requires too large a number of experiments to be applied to a real problem. In addition, it excessively cuts off the original experiment region to make the irregular experiment region symmetric when there are experimental constraints. Finally, the statistical test procedure of a conventional EVOP such as analysis of variance (ANOVA) [Neter et al., 1996] is cumbersome and its result may be ambiguous when a large number of input factors are included in the procedure. For example, we cannot clearly determine where to move the operating condition for the case that the main effects are not significant but the interaction effects are significant or that only one experiment is conducted for each experiment position due to the limitation in the number of experiments [Neter et al., 1996].

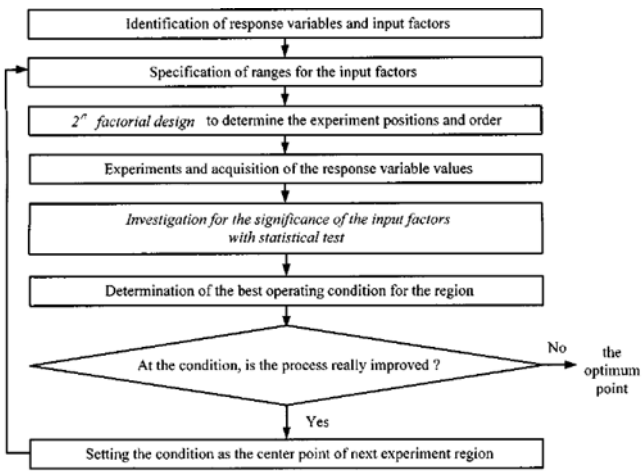

Fig. 1. Flow chart representing the procedure of the conventional EVOP.

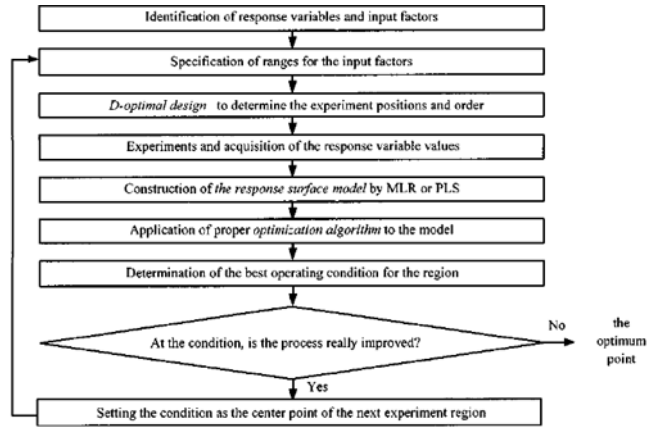

Fig. 2. Flow chart representing the procedure of the improved EVOP. In the figure, MLR and PLS represent Multiple Linear Regression and Partial Least Squares, respectively [Geladi and Kowalski, 1986].

Therefore, we adopted D-optimal design and RSM in the proposed method instead of the 2^n factorial design and statistical test procedure to solve these problems of the conventional EVOP. The procedure of the improved EVOP based on these methods is shown in Fig. 2. In the figure, the distinctive parts of the improved EVOP are denoted in italic letters. The advantages of the improved EVOP are as follows. First, the number of experiments greatly decreases especially when many input factors are used in the experiments. Second, the experiment region is taken as broadly as possible, not cutting off the original experiment region when the experiment region is irregular due to various experimental constraints. Third, the analysis results on the direction of process improvement are more reliable since we can make sure that the results of experiments are correct by the repetitive experiments in D-optimal design. Finally, the best operating condition for an experiment region can be determined more clearly by applying an appropriate optimization algorithm to the model built by RSM. These advantages will be explained in detail along with the properties of D-optimal design and RSM.

2. D-optimal Design as an Experimental Design Method

Experimental design or design of experiments (DOE) [Box and Draper, 1987] is a theory on how to arrange the experiment positions to extract important information from the results of experiments

while minimizing the number of experiments. For a process for which first principle models are not available, DOE can be used effectively to understand the input-output structure of the process. Once we know the causality of the process, we can optimally tune the input factors to improve the performance of the process. In the proposed method shown in Fig. 2, D-optimal design is used as a DOE method instead of the 2^n factorial design used in the conventional EVOP. In D-optimal design, a simple polynomial model representing the relationship between input factors and a response variable is built from the results of experiments to investigate the effects of input factors on a response variable. This model is then used to maximize the performance of a process as well as to predict a response variable value at a specific position.

The algorithm of D-optimal design is as follows [Mitchell, 1974; DuMouchel and Jones, 1994].

Step 1. Determine the proper form of the model and the number of experiments.

Step 2. Make a candidate set comprised of extreme vertices, centers of the edges, centers of faces and overall centroid of the constrained region [Piepel, 1988].

Step 3. Randomly select the initial experiment set from the candidate set so that the number of experiments in the selected subset is equal to the predetermined number of experiments.

Step 4. Interchange each experiment position in the set with a new one in the candidate set until determinant of $\mathbf{X}^T\mathbf{X}$ is maximized, where $\mathbf{X}$ is the design matrix containing the information on the arrangement of input factors within specified ranges.

When selecting experiment positions, D-optimal design employs the criterion of maximizing $\det(\mathbf{X}^T\mathbf{X})$, as described in the algorithm of D-optimal design. The following equation shows the variance-covariance matrix for the parameter vector $\mathbf{b}$ used in the model [Neter et al., 1996]:

$$\mathrm{var}\{\mathbf{b}\}=\sigma^2(\mathbf{X}^T\mathbf{X})^{-1} \quad (1)$$

From the above equation, we can see that if we maximize $\mathbf{X}^T\mathbf{X}$, the variances of the parameters are minimized and thus an accurate model can be obtained. However, $\mathbf{X}^T\mathbf{X}$ cannot be maximized since it is a matrix. Therefore, $\mathbf{X}^T\mathbf{X}$ should be made to a scalar which represents the magnitude of $\mathbf{X}^T\mathbf{X}$ and the determinant is used for this purpose in D-optimal design. That is, D-optimal design can be said to make the parameter variances as small as possible by arranging the experiment positions as broadly as possible according to the criterion, the maximization of $\det(\mathbf{X}^T\mathbf{X})$.

D-optimal design has several advantages over the factorial designs. Theoretically, we can reduce the number of experiments to the number of parameters used in the model with D-optimal design. This property of D-optimal design is effectively utilized when a large number of input factors are included in the step of experimental design. Table 1 shows the relationship between the number of input factors and the minimum number of experiments for three well-known DOE methods: 2^n factorial design, 3^n factorial design, and D-optimal design [Box and Draper, 1987; Neter et al., 1996]. Note that the D-optimal design requires the smallest number of experiments among the methods, with increasing the number of input factors.

Fig. 3 shows the difference between the factorial designs and the D-optimal design regarding the experiment region. In the figure, the use of D-optimal design does not cut off the original experiment region when the region becomes irregular due to a constraint. This property results in the derivation of the correct causal relationship between input factors and a response variable.

Table 1. The number of experiments required for the three DOE methods. As the number of input factors increases, the numbers of experiments for the two factorial designs rapidly increase

Number of input factors	2	3	4	5	6
	Number of experiments				
2^n factorial design	4	8	16	32	64
3^n factorial design	9	27	81	243	729
D-optimal design	6	10	15	21	28

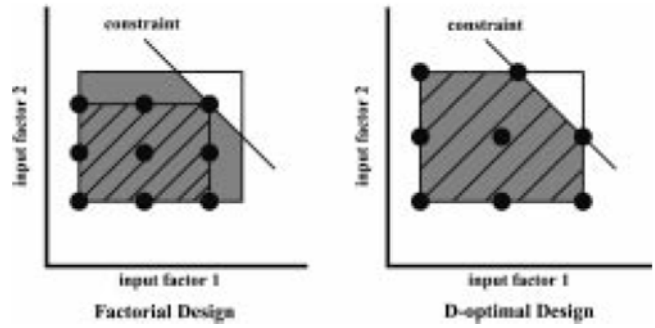

Fig. 3. Experiment regions of the factorial design and D-optimal design. Gray-colored regions mean original experiment regions.

Generally, since experiments can be repetitively conducted at a specific position in D-optimal design, it is an additional advantage that the analysis results on the direction of process improvement are more reliable than those of the factorial designs. For factorial design, usually only one experiment is conducted for each experiment position, otherwise the number of experiments increases significantly due to the symmetric property of the factorial designs. For instance, if we use 2^n factorial design, the number of experiments increases to 1024 for only 10 input factors even though only one experiment is conducted for each experiment position. On the other hand, since experiments can be repetitively conducted at a specific position in D-optimal design, we can confirm that all results for the repetitive experiments produce the same response values. Therefore, we can be convinced that the results of the experiments are correct if all results for the repetitive experiments are equal to each other, and otherwise we should carry out experiments again. This fact means that the results for the experiments and analyses obtained by D-optimal design are more reliable than by the factorial designs.

3. Response Surface Method

Response surface method is a method for building a response surface model which approximates the actual behavior of a response variable in a given experiment region by fitting response variable values obtained from designed experiments [Box and Draper, 1987]. In a conventional EVOP, the direction of process improvement is determined by analyzing the results of statistical tests with which the effects of input factors on a response variable are investigated.

Unfortunately, it happens frequently that we cannot obviously determine where to move the operating condition with the analysis results since the results are not typically clear, especially when there are a large number of input factors or when mean squared error (MSE) is small due to the small number of experiments. However, if we apply the adequate optimization algorithm to the response surface model built by RSM, the best operating condition within a given experiment region can be clearly determined.

The following response surface model described by a set of the second order polynomial equations is frequently used since it is the simplest form of nonlinear models:

$$\mathbf{y}=\mathbf{X}^T\mathbf{b} \tag{2}$$

where,

$$\mathbf{X}=\begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1n} & x_{11}^2 & x_{12}^2 & \cdots & x_{1n}^2 & x_{11}x_{12} & x_{11}x_{13} & \cdots & x_{1n-1}x_{1n} \\ 1 & x_{21} & x_{22} & \cdots & x_{2n} & x_{21}^2 & x_{22}^2 & \cdots & x_{2n}^2 & x_{21}x_{22} & x_{21}x_{23} & \cdots & x_{2n-1}x_{2n} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{m1} & x_{m2} & \cdots & x_{mn} & x_{m1}^2 & x_{m2}^2 & \cdots & x_{mn}^2 & x_{m1}x_{m2} & x_{m1}x_{m3} & \cdots & x_{mn-1}x_{mn} \end{bmatrix}^T, \tag{3}$$

$$\mathbf{b}=[\,b_{00}\ b_{10}\ b_{20}\ \cdots\ b_{n0}\ b_{11}\ b_{22}\ \cdots\ b_{nn}\ b_{12}b_{13}\cdots b_{n-1n}\,]^T, \tag{4}$$

$$\mathbf{y}=[\,y_1\ y_2\ \cdots\ y_m\,]^T \tag{5}$$

However, since the simplicity deteriorates the prediction capability of the model in a broad region, it is required to adjust the experiment region properly. If the region is too small, too many experiments should be conducted to extract the necessary information for the entire experiment region. By contrast, if the region is too large, the response surface model built in the region has poor prediction performance. Therefore, it is important to determine the optimal size of the experiment region, which guarantees both the minimization of the number of experiments and the satisfactory prediction capability of the response surface model, by using prior knowledge for a process or sensitivity analysis and so forth.

The performance of the response surface model is generally evaluated by calculating R^2 and Q^2 as follows, respectively:

$$R^2=1-\frac{SSE}{SST} \tag{6}$$

$$Q^2=1-\frac{PRESS}{SST} \tag{7}$$

where,

$$SSE(\text{Error Sum of Squares})=\sum_{i=1}^{m}(y_i-\hat{y}_i)^2, \tag{8}$$

$$SST(\text{Total Sum of Squares})=\sum_{i=1}^{m}(y_i-\bar{y})^2, \tag{9}$$

$$PRESS(\text{PREdiction Sum of Squares})=\sum_{j=1}^{s}(y_j-\hat{y}_j)^2 \tag{10}$$

R^2 represents the degree to which the variation of a response variable is explained by the response surface model and Q^2 the degree to which the variation of a response variable is predicted by the response surface model. With R^2 approaching one, the variation of a response variable explained by the response surface model increases. And with Q^2 approaching one, the prediction performance of the response surface model increases.

CASE STUDY

1. Process Description

To test the proposed method, the pulp digester benchmark model [Kayihan, 1997] was used which simulates the dynamics of a real pulp digester process. The process aims at removing the lignin from wood chips by causing reaction on the white liquor, the main constituents of which are sodium hydroxide and hydrosulfide. The entire process is composed of four sub-processes: the impregnation vessel, the cook zone, the modified continuous cooking (MCC) zone, and the extended modified continuous cooking (EMCC) zone. In Fig. 4, the manipulated variables of the process are the temperature of the mixture into the cook zone (T_{Cook}), the temperature of the trim white liqure into MCC zone (T_{MCC}), the temperature of the trim white liquor into EMCC zone (T_{EMCC}), the flowrate of the reacted white liquor extracted from the cook zone (F_{UPEX}), the flowrate of the trim white liquor into MCC zone (F_{MCC}), and the flowrate of the trim white liquor into EMCC zone (F_{EMCC}). These six variables are used as input factors for the experiments. The Kappa number defined by the following expression represents the lignin content in wood chips and is the response variable that should be minimized in the process:

$$\text{Kappa number}=654\frac{\text{the lignin mass}}{\text{the total solid mass}} \tag{11}$$

Since both the conventional EVOP and the improved EVOP are the methods which are directly applied to a real process itself, all the experiments in this study are regarded as actual experiments although they are simulations. On the assumption that the first principle models cannot be built for the process, we apply both the con-

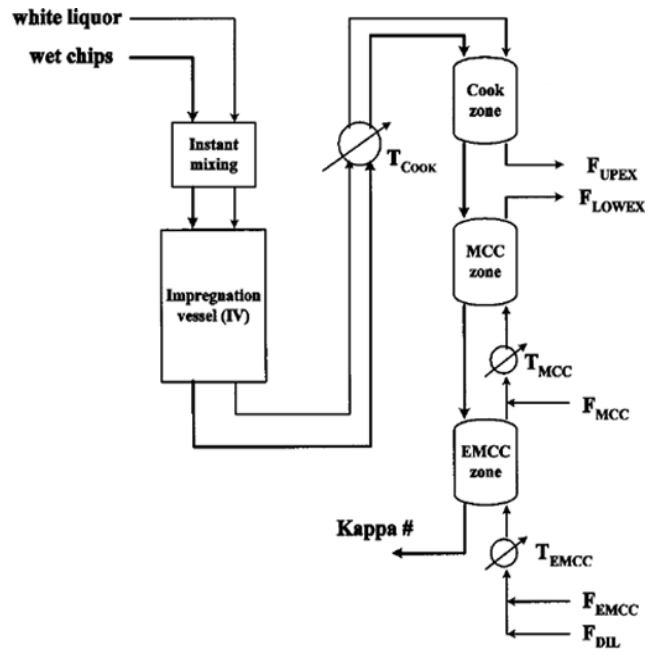

Fig. 4. Process diagram of the pulp digester benchmark model. MCC and EMCC stand for modified continuous cooking and extended modified continuous cooking, respectively.

Table 2. ANOVA table for the first phase (MSE[1]=0.0214)

Term	Degree of freedom[2]	Sum of squares[3]	Mean square[4]	F statistics[5]	P-value[6]
T_{Cook}	1	407.3722	407.3722	19048.5475	0.0001
T_{MCC}	1	402.1474	402.1474	18804.2388	0.0001
T_{EMCC}	1	389.3918	389.3918	18207.7928	0.0001
F_{UPEX}	1	668.9536	668.9536	31279.9839	0.0001
F_{MCC}	1	618.5026	618.5026	28920.9164	0.0001
F_{EMCC}	1	1089.6477	1089.6477	50951.4602	0.0001
$T_{Cook}T_{MCC}$	1	2.1118	2.1118	98.7482	0.0001
$T_{Cook}T_{EMCC}$	1	1.6958	1.6958	79.2929	0.0001
$T_{Cook}F_{UPEX}$	1	1.5778	1.5778	73.7781	0.0001
$T_{Cook}F_{MCC}$	1	3.1690	3.1690	148.1801	0.0001
$T_{Cook}F_{EMCC}$	1	4.4849	4.4849	209.7128	0.0001
$T_{MCC}T_{EMCC}$	1	4.7361	4.7361	221.4590	0.0001
$T_{MCC}F_{UPEX}$	1	0.7070	0.7070	33.0594	0.0001
$T_{MCC}F_{MCC}$	1	1.2893	1.2893	60.2886	0.0001
$T_{MCC}F_{EMCC}$	1	0.0245	0.0245	1.1447	0.2908
$T_{EMCC}F_{UPEX}$	1	1.0912	1.0912	51.0223	0.0001
$T_{EMCC}F_{MCC}$	1	0.6884	0.6884	32.1904	0.0001
$T_{EMCC}F_{EMCC}$	1	3.0772	3.0772	143.8873	0.0001
$F_{UPEX}F_{MCC}$	1	9.2102	9.2102	430.6668	0.0001
$F_{UPEX}F_{EMCC}$	1	11.0104	11.0104	514.8401	0.0001
$F_{MCC}F_{EMCC}$	1	13.7160	13.7160	641.3545	0.0001
Error	42	0.8982	0.0214		
Total	63	3635.5031			

[1]MSE=SSE/(Degree of freedom for error)

[2]Degree of freedom (for each term)=number of levels for the corresponding term−1=2−1=1. Degree of freedom for error and total data is defined as "m−1−number of terms used" and "m−1".

[3]Sum of Squares=total sum of (mean for the corresponding term−overall mean)^2

[4]Mean Square=Sum of Squares/(Degree of freedom)

[5]F statistics=Mean Square/MSE

[6]P-value=right portion of the F statistics in the corresponding F-distribution, that is, the probability that the term is not significant compared with error

ventional EVOP and the improved EVOP to the process to find the optimal set of the input factors at which Kappa number is minimized. Note that we assume that noises and disturbances, which may occur in the real process, are filtered out to observe the effects of the input factors apparently since there are no noises and disturbances in the simulated values.

2. Application of the Conventional EVOP

At first, we applied the conventional EVOP to the pulp digester benchmark model. According to the procedure shown in Fig. 1, we designed experiments with a 2^n factorial design to determine the experiment positions after specifying the appropriate ranges of the input factors. Since there are the six input factors, 2^6 (=64) experiments were run for each phase. A phase means one implementation of the whole procedure shown in Fig. 1 or Fig. 2 and we completed three phases to find the optimal operating condition. If we use a 3^n factorial design instead of the 2^n factorial design, the number of experiments increases to 3^6 (=729). In addition, it should be noticed that the better operating condition which may exist between the upper and lower levels of the input factors cannot be detected with the 2^n factorial design.

After 64 experiments were completed based on the results of the 2^n factorial design, we statistically tested the significance of various effects using ANOVA to find the direction of process improvement (Table 2). At the first phase, all the terms except the $T_{MCC}F_{EMCC}$ term shown in Table 2 were significant at 99% confidence level since their p-values (right portion of the F statistics in the corresponding F-distribution) were smaller than 0.01. The result means that if we change the value of each input factor from one level to the other level, the Kappa number changes in a statistically significant manner. The direction of change in the Kappa number can be inferred from the knowledge on the process. For example, the Kappa number is expected to decrease because of the increase in the reaction rate if we increase T_{Cook}. Based on the knowledge and ANOVA results, we concluded that all of the main effect terms (T_{Cook}, T_{MCC}, T_{EMCC}, F_{UPEX}, F_{MCC}, and F_{EMCC}) had negative and significant effects on the Kappa number. Therefore, we set all the input factors at the upper levels within their ranges to reduce the Kappa number. The significance of the interaction effect terms (the terms except for the main effect terms in Table 2) implies that the effect of each input factor is not additive and the process has nonlinearity in this region.

However, the mean square values for the interaction effect terms are so small compared with those for the main effect terms that the interaction effects may be ignored. Moreover, it should be noted that we may draw an erroneous conclusion from the ANOVA results since the significance of all the terms was derived from a small mean squared error value caused by the small number of experiments.

Since the Kappa number was reduced from 29.7091 to 15.0705 when all of the input factors were set at the upper level in the first phase, we entered the second phase of the conventional EVOP. At the second and third phases, experimental constraints described by Eqs. (12) and (13) were introduced to the experimental design step on the assumption that experiments at specific points cannot be conducted due to environmental or economical reasons:

$$0.00185T_{Cook}+F_{UPEX}\leq 0.893 \quad (12)$$

$$0.419F_{UPEX}+F_{EMCC}\leq 0.0528 \quad (13)$$

Eq. (12) means that if the temperature of the mixture into the cook zone is too high, we must reduce the flowrate of the reacted white liquor extracted from the cook zone. This constraint reflects the environmental restriction that hot wastewater should not be discharged into nature in a large amount. Eq. (13) takes into account the operating costs. That is, if the reacted white liquor is extracted sufficiently from the cook zone, the flowrate of the trim white liquor into EMCC zone should be reduced since most of the lignin in the wood chips is expected to be already removed at the cook zone. In addition to these constraints, the upper limits to which the input factors can be adjusted were specified in the third phase since we cannot increase the input factors infinitely. Considering these constraints, we determined the ranges of the input factors as shown in Table 3. As for the second phase, since all of the terms were significant at 99% confidence level, all the input factors were again set at the upper levels within their ranges despite some ambiguity in the ANOVA results.

Because the Kappa number again decreased to 4.4173 at the operating condition found in the second phase, we again entered the third phase. After performing the same procedure as the previous phases,

Table 3. The determined ranges of the input factors for the conventional EVOP

	Input factors						Response variable
	T_{Cook} (K)	T_{MCC} (K)	T_{EMCC} (K)	F_{UPEX} (m^3/min)	F_{MCC} (m^3/min)	F_{EMCC} (m^3/min)	Kappa number
Nominal operating condition	425	420	415	0.09	0.01	0.01	29.7091
Upper limit	435	430	432	0.1	0.02	0.017	
1st phase	±3	±2	±5	±0.003	±0.002	±0.0017	
2nd phase	±5	±4	±10	±0.0031	±0.004	±0.0026	
3rd phase	±5	±4	±4.3	±0.0096	±0.0046	±0.0009	

Table 4. ANOVA table for the third phase (MSE=0.9208)

Term	Degree of freedom	Sum of squares	Mean square	F statistics	P-value
T_{Cook}	1	10.5279	10.5279	11.4335	0.0016
T_{MCC}	1	303.1830	303.1830	329.2608	0.0001
T_{EMCC}	1	47.5748	47.5748	51.6669	0.0001
F_{UPEX}	1	4196.3383	4196.3383	4557.2802	0.0001
F_{MCC}	1	1834.0513	1834.0513	1991.8046	0.0001
F_{EMCC}	1	174.2941	174.2941	189.2858	0.0001
$T_{Cook}T_{MCC}$	1	6.7279	6.7279	7.3066	0.0099
$T_{Cook}T_{EMCC}$	1	1.3117	1.3117	1.4245	0.2394
$T_{Cook}F_{UPEX}$	1	18.3160	18.3160	19.8915	0.0001
$T_{Cook}F_{MCC}$	1	0.0018	0.0018	0.0019	0.9651
$T_{Cook}F_{EMCC}$	1	0.0739	0.0739	0.0802	0.7784
$T_{MCC}T_{EMCC}$	1	17.2698	17.2698	18.7553	0.0001
$T_{MCC}F_{UPEX}$	1	11.9633	11.9633	12.9923	0.0008
$T_{MCC}F_{MCC}$	1	4.1417	4.1417	4.4980	0.0399
$T_{MCC}F_{EMCC}$	1	0.4351	0.4351	0.4726	0.4956
$T_{EMCC}F_{UPEX}$	1	0.0316	0.0316	0.0343	0.8539
$T_{EMCC}F_{MCC}$	1	0.6475	0.6475	0.7032	0.4065
$T_{EMCC}F_{EMCC}$	1	0.1139	0.1139	0.1237	0.7268
$F_{UPEX}F_{MCC}$	1	876.4486	876.4486	951.8351	0.0001
$F_{UPEX}F_{EMCC}$	1	62.4882	62.4882	67.8631	0.0001
$F_{MCC}F_{EMCC}$	1	28.5015	28.5015	30.9530	0.0001
Error	42	38.6736	0.9208		
Total	63	7633.1155			

Table 5. The determined ranges of the input factors for the improved EVOP

	Input factors						Response variable
	T_{Cook} (K)	T_{MCC} (K)	T_{EMCC} (K)	F_{UPEX} (m^3/min)	F_{MCC} (m^3/min)	F_{EMCC} (m^3/min)	Kappa number
Nominal operating condition	425	420	415	0.09	0.0	0.01	29.7091
Upper limit	435	430	432	0.1	0.02	0.017	
1st phase	±3	±2	±5	±0.003	±0.002	±0.0017	
2nd phase	±6	±4	±10	±0.005	±0.004	±0.0034	
3rd phase	±5.8	±4.1	±4.3	±0.013	±0.005	±0.0019	

we obtained the ANOVA table shown in Table 4. In this table, the eight interaction effect terms ($T_{Cook}T_{EMCC}$, $T_{Cook}F_{MCC}$, $T_{Cook}F_{EMCC}$, $T_{MCC}F_{MCC}$, $T_{MCC}F_{EMCC}$, $T_{EMCC}F_{UPEX}$, $T_{EMCC}F_{MCC}$, and $T_{EMCC}F_{EMCC}$) are not significant at 99% confidence level as their p-values are larger than 0.01. Moreover, the mean square values for T_{Cook} and T_{EMCC} terms are relatively small compared to those for the other main effect terms. Therefore, the Kappa number may not change significantly when these two variables are changed from one level to the other level. However, we set all the input factors at the upper levels since all the main effect terms were statistically significant at 99% confidence level, and the Kappa number was again reduced to 2.8448.

Table 6. The experiment set selected by the D-optimal design and experiment results

Exp. no.	T_{Cook} (K)	T_{MCC} (K)	T_{EMCC} (K)	F_{UPEX} (m^3/min)	F_{MCC} (m^3/min)	F_{EMCC} (m^3/min)	Kappa number
1	422	418	410	0.0875	0.008	0.0083	50.4072
2	428	418	420	0.0875	0.008	0.0083	39.2747
3	422	422	420	0.0875	0.008	0.0083	40.3719
4	428	418	410	0.0925	0.008	0.0083	35.3751
5	422	422	410	0.0925	0.008	0.0083	37.1424
6	422	418	420	0.0925	0.008	0.0083	37.2859
7	422	418	420	0.0875	0.012	0.0083	37.0905
8	422	422	410	0.0875	0.008	0.0117	34.7424
9	422	418	420	0.0875	0.008	0.0117	33.7384
10	422	418	410	0.0925	0.008	0.0117	34.2205
11	422	422	420	0.0925	0.008	0.0117	23.4967
12	428	422	420	0.0875	0.012	0.0117	19.5513
13	428	422	410	0.0925	0.012	0.0117	18.8359
14	422	418	410	0.0925	0.011	0.0083	37.9273
15	422	418	413	0.0925	0.012	0.0083	34.4112
16	422	422	420	0.0925	0.012	0.0094	23.132
17	422	419	410	0.0875	0.012	0.0117	32.1861
18	428	418	410	0.0875	0.012	0.0106	31.9846
19	428	418	410	0.0892	0.008	0.0117	32.746
20	428	418	420	0.0925	0.009	0.0117	21.3095
21	428	422	410	0.0875	0.009	0.0083	36.3788
22	428	422	420	0.0925	0.008	0.0106	21.3935
23	428	422	420	0.0925	0.011	0.0083	22.8632
24	428	421	410	0.0875	0.008	0.0117	31.8598
25	428	419	420	0.0925	0.012	0.0083	23.9994
26	424	418	420	0.0925	0.012	0.0117	21.5992
27	424	422	410	0.0875	0.012	0.0083	34.0915
28	422	420	415	0.09	0.01	0.01	32.511
29	425	420	410	0.09	0.01	0.01	32.0554
30	425	420	415	0.0875	0.01	0.01	33.1293
31	425	420	415	0.09	0.01	0.01	29.7091
32	425	420	415	0.09	0.01	0.01	29.7091
33	425	420	415	0.09	0.01	0.01	29.7091

3. Application of the Improved EVOP

We also applied the improved EVOP which is the proposed method to the pulp digester benchmark model according to the procedure shown in Fig. 2. As the first step of the D-optimal design, the form of the model and the number of experiments were determined in advance as a quadratic model and 33, respectively. Since the quadratic model requires 28 parameters concerning six input factors, at least 28 experiments should be conducted. Thus, we determined the number of experiments per phase as 33, considering repetitive experiments at the center point. Unlike the factorial design where only one experiment is usually conducted for each experiment position, these repetitive experiments at a specific position in D-optimal design allow us to know whether the same results are obtained or not at the position. Correspondingly, the results of experiments and analyses using the D-optimal design have more reliability than using the factorial design. Table 6 shows the result of the D-optimal design and Kappa numbers obtained at each experiment position for the first phase. Based on the experiment results, we built the quadratic response surface model regarding the six input factors and one response variable which is the Kappa number. After performing the initial regression, we discarded seven terms in the model that have very small parameter values, and then rebuilt the qua-

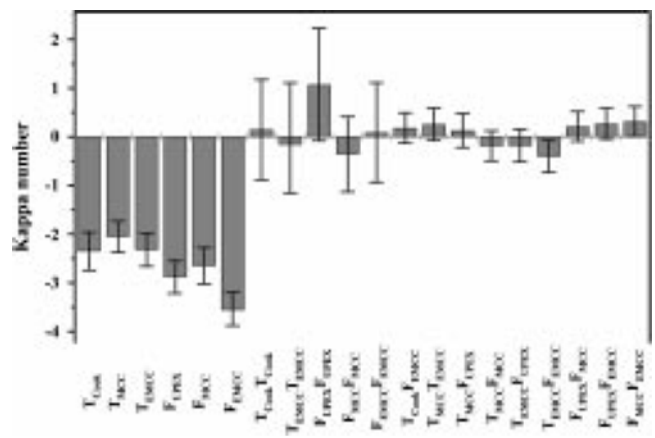

Fig. 5. Column plot for the coefficients of the response surface model built at the first phase. The coefficient values are mean-centered and scaled to unit variance.

Table 7. The best operating conditions and resultant kappa numbers for each phase

	Input factors						Response variable
	T_{Cook} (K)	T_{MCC} (K)	T_{EMCC} (K)	F_{UPEX} (m^3/min)	F_{MCC} (m^3/min)	F_{EMCC} (m^3/min)	Kappa number
1st phase	428	422	420	0.0924	0.012	0.0117	15.1897
2nd phase	429	426	428	0.0874	0.0154	0.0151	6.2836
3rd phase	429	430	429	0.0904	0.0157	0.0166	3.0394

dratic model with 20 remaining important terms. The parameter values of the response surface model at the first phase are shown in Fig. 5 in the form of mean-centered and unit-variance-scaled coefficient values. This figure shows that the parameter values of the first order terms are large compared to those of the second order terms, which means that the nonlinearity in this experiment region is not so severe. Besides, since all the parameter values of the first order terms have negatives, if the input factors are increased in the region, the Kappa number is expected to decrease. These results agree with real phenomena. The fact that R^2 is 0.9952 and Q^2 0.8770 implies that the response surface model at the first phase explains the variation of the experiment data quite accurately and well predicts the behavior of the Kappa number in the experiment region.

To clearly find the values of the input factors at which the Kappa number was minimized in the region, we applied the Nelder-Mead simplex method [Nelder and Mead, 1965] to the response surface model, and obtained the optimal point in the experiment region (Table 7). When we set the values as the operating condition of the pulp digester benchmark model, the Kappa number decreased from 29.7091 to 15.1897 as shown in Fig. 6.

The procedure implemented at the first phase was also repeated at another two phases, and the constraints used at the second and third phases of the conventional EVOP were also considered at these phases. The determined ranges of the input factors at these phases are shown in Table 5. The parameter values of the response surface model built at the third phase are shown in Fig. 7. In this model, only 19 important terms were used for the regression, and the R^2 and Q^2 of this model were 0.9781 and 0.8322, respectively. Note that the coefficients of the response surface model in the experiment region are different from those of the response surface model constructed at the first phase because of the constraints and nonlinearity. When the best point found in the region by applying the optimization algorithm to the response surface model was set as the operating condition of the pulp digester benchmark model, the Kappa number finally decreased to 3.0394 as shown in Fig. 8.

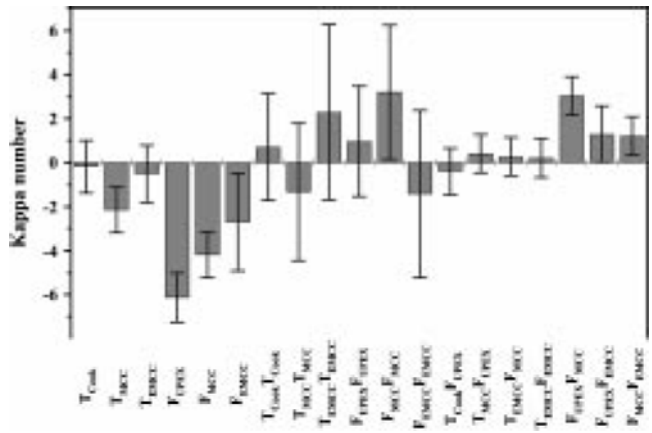

Fig. 7. Column plot for the coefficients of the response surface model built at the third phase. The coefficient values are mean-centered and scaled to unit variance.

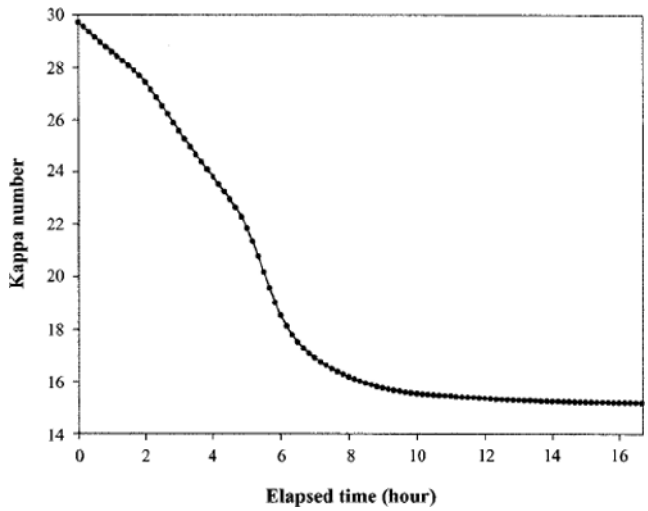

Fig. 6. Decrease in the kappa number at the best operating condition of the first phase. The kappa number rapidly decreases during the 8 hours from the start, and then reach the new steady state.

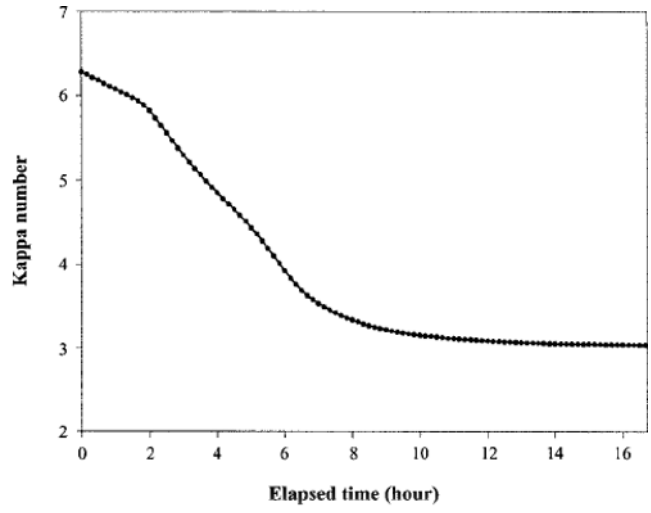

Fig. 8. Decrease in the kappa number at the best operating condition of the third phase. The kappa number remarkably decreases during the 8 hours from the start, and then reach the new steady state.

Table 8. Comparison between the conventional EVOP and the improved EVOP

	No. of exp. (For one phase)	Preservation of original exp. region	Reliability of results	Detection of the optimum
Improved EVOP	33	Yes	High	Clear
Conventional EVOP	64	No	Low	Maybe ambiguous

4. Comparison of the Conventional EVOP and the Improved EVOP

We have applied both the conventional EVOP and the improved EVOP to the pulp digester benchmark model to find the optimal operating condition on the assumption that the characteristics of the process were not known. We can summarize the comparison results as follows (Table 8). First, the number of experiments for each phase was 64 for the conventional EVOP and 33 for the improved EVOP. Second, when there were experimental constraints, the original experiment region was maintained for the improved EVOP while it shrank for the conventional EVOP. If we compare Tables 3 and 5, we can see that the ranges of T_{Cook}, F_{UPEX} and F_{EMCC} for the improved EVOP are larger than those for the conventional EVOP. Third, in the case of the improved EVOP, we can have confidence in the results of experiments and analyses by repeating experiments at a specific position. Finally, the best operating conditions for each experiment region were obtained more clearly by virtue of the response surface model adopted in the improved EVOP.

Although a smaller Kappa number, 2.8448, was obtained for the conventional EVOP compared with 3.0394 for the improved EVOP, the two values do not show significant differences. Moreover, it can be surely said that the improved EVOP is a much more efficient method than the conventional EVOP from the viewpoint of the previously mentioned advantages.

CONCLUSIONS

In this paper, an improved evolutionary operation (EVOP) based on D-optimal design and response surface method (RSM) was proposed. To test its performance and superiority to a conventional EVOP, we applied both the conventional EVOP and the improved EVOP to the pulp digester benchmark model. As a result, the improved EVOP showed comparable performance with advantages on the number of experiments, preservation of original experiment region, reliability in the results of experiments and analyses, and determination of the best operating condition for each experiment region. These benefits are generated by the properties of D-optimal design and RSM and make the proposed method more applicable in practical problems than the conventional EVOP.

Although we enhanced the performance of the conventional EVOP by modifying its several properties, we have to keep in mind that the key point for the successful application of the proposed method to the real industry is the harmonious collaboration among all the participants. In addition, the selection of the appropriate process, pre-experiment on the pilot plant and education for industrial operators should precede the real application of the method.

ACKNOWLEDGEMENT

This paper was supported by the Brain Korea 21 Project.

NOMENCLATURE

b : parameter vector of a quadratic response surface model, with the size of $\left(1+2n+\frac{n(n-1)}{2}\right)$ by 1

b_{ij} : the coefficient for the cross product term of ith input factor and jth input factor in the response surface model

F_{EMCC} : flowrate of the trim white liquor into the extended modified continuous cooking zone [m^3/min]

F_{MCC} : flowrate of the trim white liquor into the modified continuous cooking zone [m^3/min]

F_{UPEX} : flowrate of the reacted white liquor extracted from the cook zone [m^3/min]

Kappa number : mass fraction of the lignin content in wood chips [-]

m : number of experiment data used in modeling

MSE : mean squared error [=SSE/(m−1−number of terms used)]

n : number of input factors

PRESS : prediction sum of squares

Q^2 : the degree to which the variation of a response variable is predicted by the response surface model [-]

R^2 : the degree to which the variation of a response variable is explained by the response surface model [-]

s : number of experiment data not used in modeling

SSE : error sum of squares

SST : total sum of squares

T_{Cook} : temperature of the mixture into the cook zone [K]

T_{EMCC} : temperature of the trim white liquor into the extended modified continuous cooking zone [K]

T_{MCC} : temperature of the trim white liquor into the modified continuous cooking zone [K]

X : design matrix with the size of $\left(1+2n+\frac{n(n-1)}{2}\right)$ by m containing the information of the experiment positions

x_{kl} : kth experiment position of lth input factor

y : vector of a response variable with the size of m by 1

y_i : ith real response value used in modeling

y_j : jth real response value not used in modeling

$\bar{y}$: overall mean of the response values

$\hat{y}_i$: ith response value fitted by the response surface model

$\hat{y}_j$: jth response value predicted by the response surface model

Greek Letter

σ^2 : variance of model error

Subscripts

Cook : cook zone

DIL : dilution water
EMCC : extended modified continuous cooking zone
i : index of response values used in modeling
j : index of response values not used in modeling
k : index of experiment positions
l : index of input factors
LOWEX : lower extract from the modified continuous cooking zone
MCC : modified continuous cooking zone
UPEX : upper extract from the cook zone

REFERENCES

Banerjee, R. and Bhattacharyya, B. C., "Evolutionary Operation to Optimize Three-Dimensional Biological Experiments," *Biotechnol. Bioeng.*, **41**, 67 (1993).

Box, G. E. P., "Evolutionary Operation: A Method for Increasing Industrial Productivity," *Appl. Statist.*, **4**, 3 (1957).

Box, G. E. P. and Draper, N. R., "Empirical Model-Building and Response Surfaces," John Wiley & Sons, New york, USA (1987).

Box, G. E. P. and Wilson, K. B., "On the Experimental Attainment of Optimum Conditions," *J. Roy. Statist. Soc.*, **B13**, 1, discussion 38 (1951).

Choi, S. and Manousiouthakis, V., "Global Optimization Methods for Chemical Process Design: Deterministic and Stochastic Approaches," *Korean J. Chem. Eng.*, **19**, 227 (2002).

Choi, Y., Kwon, T. and Yeo, Y., "Optimization of the Sulfolane Extraction Plant Based on Modeling and Simulation," *Korean J. Chem. Eng.*, **17**, 712 (2000).

DuMouchel, W. and Jones, B., "A Simple Bayesian Modification of D-Optimal Designs to Reduce Dependence on an Assumed Model," *Technometrics*, **36**, 37 (1994).

Galil, Z. and Kiefer, J., "Time- and Space-Saving Computer Methods, Related to Mitchells DETMAX, for Finding D-Optimum Designs," *Technometrics*, **21**, 301 (1980).

Geladi, P. and Kowalski, B. R., "Partial Least-Squares Regression: A Tutorial," *Anal. Chim. Acta*, **185**, 1 (1986).

Hunter, W. G. and Kittrell, J. R., "Evolutionary Operation: A Review," *Technometrics*, **8**, 389 (1966).

Janson, J., "Optimization of Large-Scale Chromatography of Proteins," *Korean J. Chem. Eng.*, **18**, 149 (2001).

Johnson, M. E. and Nachtsheim, C. J., "Some Guidelines for Constructing Exact D-Optimal Designs on Convex Design Spaces," *Technometrics*, **25**, 271 (1983).

Kayihan, F., "Continuous Digester Benchmark Model v1.0," IETek, Washington, USA (1997).

Lee, J. and Lim, H., "Development of Nonsingular Optimization Algorithm and Its Application to Chemical Engineering Systems," *Korean J. Chem. Eng.*, **16**, 118 (1999).

Lee, S., Lee, M., Han, C. and Chang, K., "Optimization of a Utility Plant Operation Based on a Hierarchical Method," *HWAHAK KONGHAK*, **36**, 415 (1998).

Lee, S., Lee, M., Han, C. and Chang, K., "Optimization of a Utility Plant for Multi-period Operations," *HWAHAK KONGHAK*, **36**, 422 (1998).

Mitchell, T. J., "An Algorithm for the Construction of D-Optimal Experimental Designs," *Technometrics*, **16**, 203 (1974).

Nelder, J. A. and Mead, R., "A Simplex Method for Function Minimization," *Computer J.*, **7**, 308 (1965).

Neter, J., Kutner, M. H., Nachtsheim, C. J. and Wasserman, W., "Applied Linear Statistical Models," 4th ed., IRWIN, New York, USA (1996).

Park, J., Ha, D. and Kim, M., "Modified Response Surface Methodology (MRSM) for Phase Equilibrium-Theoretical Background," *Korean J. Chem. Eng.*, **13**, 115 (1996).

Piepel, G. F., "Programs for Generating Extreme Vertices and Centroids of Linearly Constrained Experimental Regions," *J. Qual. Technol.*, **20**, 125 (1988).

Powell, M. J. D., "An Efficient Method for Finding the Minimum of a Function of Several Variables without Calculating Derivatives," *Computer J.*, **7**, 155 (1964).

Saad, E. E., "An EVOP Design for Process Optimization," *Ceram. Eng. Sci. Proc.*, **15**, 17 (1994).

Tunga, R., Banerjee, R. and Bhattacharyya, B. C., "Optimization of n Variable Biological Experiments by Evolutionary Operation-Factorial Design Technique," *J. Biosci. Bioeng.*, **87**, 224 (1999).

Yi, H. and Han, C., "The Integration of Complete Replanning and Rule-Based Repairing for Optimal Operation of Utility Plants," *Korean J. Chem. Eng.*, **18**, 442 (2001).