Product and Process Development Using Artificial Neural-Network Model and Information Analysis

Junghui Chen, David Shan Hill Wong, and Shi-Shang Jang
Dept. of Chemical Engineering, National Tsing-Hua University, Hsin-Chu, Taiwan 30034, R.O.C.

Seng-Lu Yang
China Glaze Co., Ltd., Hsin-Chu, Taiwan, R.O.C.

An experimental design scheme proposed for process and product development integrates the artificial neural network, random-search algorithm, fuzzy classification, and information theory. An initial batch of experimental data is first collected to construct a neural-network model. Random search generates a number of candidates for the next batch of experiments. A fuzzy classification algorithm is used to find the cluster centers of these candidates. An information free energy index is defined to balance the need for better classification and the relevance of each class in optimization. New experiments are performed at these cluster centers to validate the model. The procedure is repeated until an optimal solution is reached. Case studies using a mathematical model and a real industrial pigment-blending project illustrate the abilities of this method to locate multiple optima and handle multivariable experimental design.

Introduction

In the competitive market, speed product or process development is the key to success due to shorter product life cycles. If the new product cannot be made in time to meet the needs on the market, it would be outdated or even no longer wanted. It is important to accelerate the process design procedure without sacrificing product quality and production cost. It is also important to have quality-improvement strategies start at the *design stage* rather than the *product inspection stage*. In many industries, such as specialty chemicals, ceramic and composite materials, finding recipes, and designing new processes are basically empirical. Getting experimental data, if not difficult, is time-consuming and costly. When there is an abundance of data, an experienced engineer is needed to sort through them and detect opportunities of improvement. Traditionally, a systematic methodology that includes statistical data analysis and decision making is known as *experimental design* (Box and Draper, 1987; Taguchi, 1986; Lochner and Matar, 1990). On the other hand, product and process development are regarded as *learning* experiences that have been the focus of many artificial intelligence researchers. For example, Fukunaga (1990) described the process of *classifier design* or *statistical pattern recognition* in a series of iterative steps: data gathering, registry, analysis, classifier design, and testing. The logic does not differ from that of *experimental design*, except that the tools employed are more suitable for problems with high dimensionality and nonlinearity. Saraiva and Stephanopoulos (1992) demonstrated that with existing plant data, one class of machine learning approaches, top-down induction of decision trees, can be used to explore process-improvement opportunities.

In this research, a new experimental design scheme that uses the artificial neural network (ANN), random search, fuzzy classification, and information theory is proposed. It can find out the potentially available knowledge of the process and reduce the time taken when the experimental study is undertaken. This approach is quite different from that of Lin et al. (1995), whose approach for experimental design combines simulated annealing and ANN modeling methods. This work replaces their stochastic elements using a deterministic method derived from information theory, because the stochastic search is lack of strong information for decision-makers. The major advantages of the proposed method are its abilities to cope with multivariables, precisely determine the number and location of future test experiments, and to locate multiple optima. Various components of the proce-

Correspondence concerning this article should be addressed to S.-S. Jang.

dure are elaborated in the following five sections. Two case studies are presented in the seventh section. Finally, conclusions are made.

Neural-Network Model

A model is a summary of our knowledge about performance–variable relations. There are two common ways to construct a model. The first option is to derive a model from physical principles. Unfortunately, due to the complexity of the processes involved, construction of an applicable physical model is time-consuming, expensive, and even impossible in many cases. Alternatively, experimental data or historical plant data are used to build statistical or black-box models. Artificial neural networks are known to be a powerful tool to approximate complex multivariable functions (Hornik et al., 1989, 1990). In our first step, a feedforward artificial neural network of currently available data is constructed using supervised learning.

The network in this research is composed of three layers (Figure 1). The input variables of the network are the design variables $\boldsymbol{x}^p = [x_1^p, \ldots, x_{N_i}^p]$. The hidden layer is composed of N_i neurons, whose output is given by

$$h_j(\boldsymbol{x}^p) = \sum_{i=1}^{N_i} w_{ji}^h x_i^p + b_j^h, \qquad j = 1, 2, \ldots, N_h. \tag{1}$$

The outputs of the network represent the predicted controlled or response variables, $\boldsymbol{y}^p = [y_1^p, \ldots, y_{N_o}^p]$, with N_h neurons in the hidden layer. The output is given by

$$\hat{y}_k(\boldsymbol{x}^p) = \sum_{j=1}^{N_h} w_{kj}^o z[h_j(\boldsymbol{x}^p)] + b_k^o, \qquad k = 1, 2, \ldots, N_o, \tag{2}$$

where N_o is the dimensions of the output vectors, the input and the output component of the pth data pair are defined by $\{\boldsymbol{x}^p, \boldsymbol{y}^p\}$; w_{kj}^o and w_{ji}^h are weights between the output and the hidden layers and weights between the hidden and the input layers, respectively; b_k^o and b_j^h are biases in the output layers and hidden layers respectively; and z is the output in the hidden layer. The hyperbolic tangent activation function is used (Kalman and Kwasny, 1992). The sum square error, E, which represents the error between the predicted and targeted values is employed to evaluate the ability of the network,

$$E = \frac{1}{P} \sum_{p=1}^{P} \sum_{k=1}^{N_o} [y_k(\boldsymbol{x}^p) - \hat{y}_k(\boldsymbol{x}^p)]^2, \tag{3}$$

where P is the number of experimental data. The pseudo-Gauss–Newton method (Hertz et al., 1991; Gorodkin et al., 1993) is used for training. Due to the small number of training data, a statistical technique called the leave-one-out (LOO) cross-validation scheme is used (Allen, 1974).

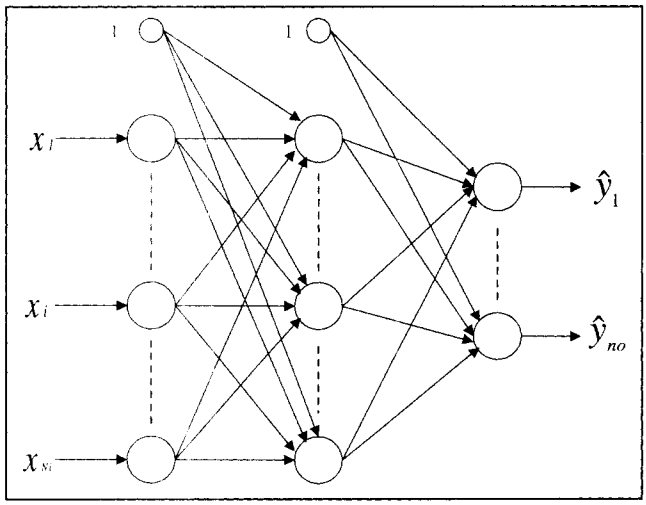

Figure 1. Architecture of a feedforward neural network.

Region Optimal Search

To demonstrate the applicability of a model, we need to extract features from the model and validate them. In product and process development, the feature of interest is the optimal operating condition. Multiple local optima are frequently encountered. It is often necessary to rate alternative local optima-based secondary objectives such as robustness, safety, and the like. Therefore, a nongradient-based search procedure should be used. An existing model is most trustworthy around the experimental points. They should therefore be used as the starting points of our search. The entire procedure is described as follows:

Step 1. Use existing experimental points as starting points. Define a local search region as a hypersphere around these experimental points with volume equal to total search space divided by the number of starting points (Figure 2a and 2b).

Step 2. Generate a set of N_r random points at each starting point and evaluate the objective function at these points.

Step 3. Extract the best N_s points. Define a local search region as a hypersphere with a volume equal to the total search space divided by N_s. Reset N_s as new starting points.

Step 4. Repeat steps 2 and 3 until the average performance of the best N_s points has no significant change (Figure 2c).

It should be pointed out that other nongradient-based procedures that are capable of locating multiple local minima, and retaining a near-optimum population such as a genetic algorithm could be used for this step. The key issue is what we do with this population. It is impossible to perform experiments at all the points. A clustering technique is used here to select the most representative candidate points. Experiments will be performed only at the clustering centers.

Fuzzy Classification

A clustering technique is used here to select the most representative candidate points for the best performance. The purpose of the clustering process is to distill a certain number of homogeneous clusters or classes from a large data set and to classify a concise representation of the individual local optimal behavior. Experiments will then be performed only at the clustering centers. Data points of all-or-none crisp classes are not fitted, because in reality each point may belong to more than one cluster. The degree of its membership characterizes the point as being, to a greater or lesser degree,

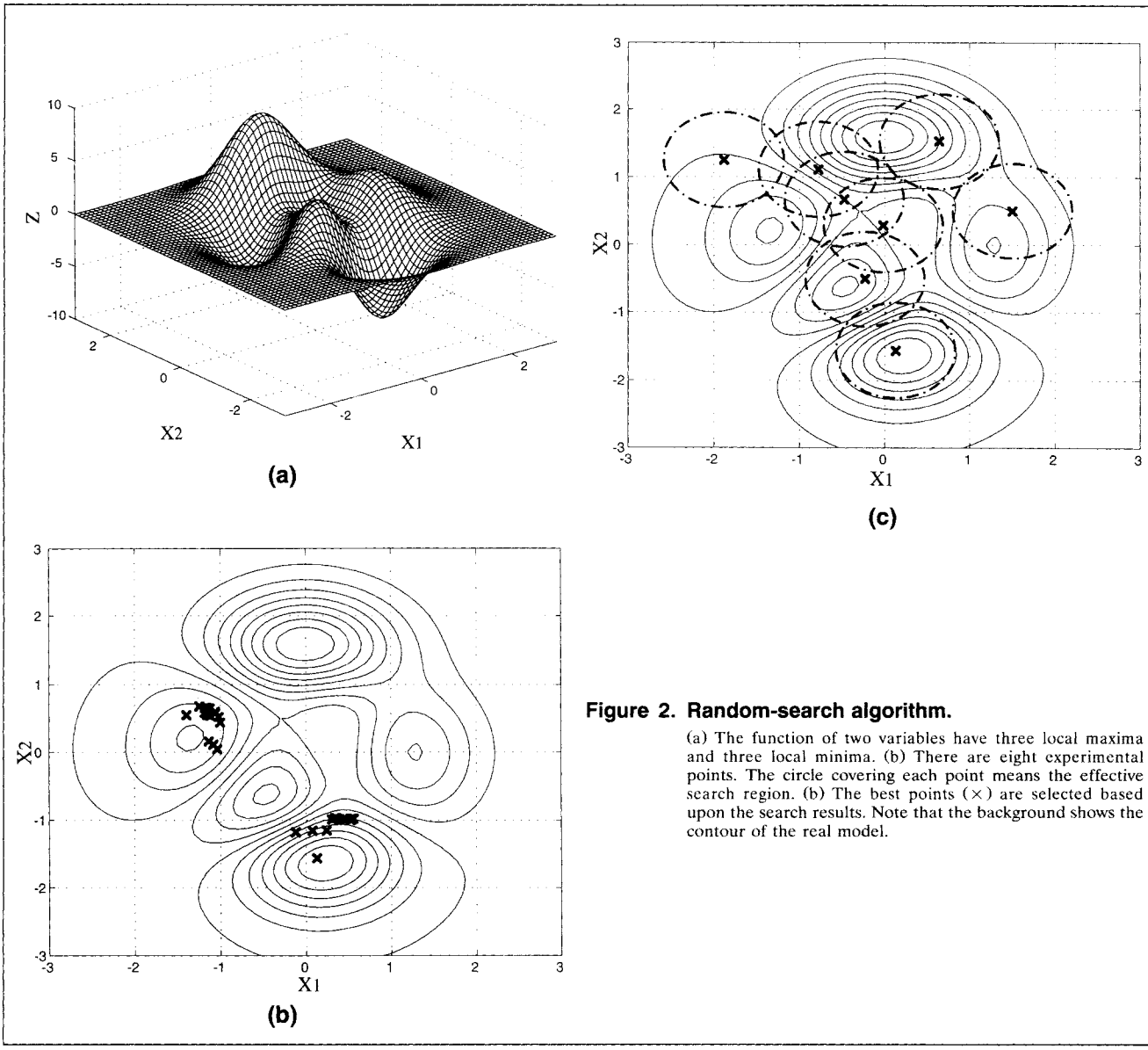

Figure 2. Random-search algorithm.

(a) The function of two variables have three local maxima and three local minima. (b) There are eight experimental points. The circle covering each point means the effective search region. (b) The best points (×) are selected based upon the search results. Note that the background shows the contour of the real model.

a member of the appropriate cluster. A fuzzy classification technique is selected to divide these candidates into several groups or clusters.

The classification algorithm we use is an unsupervised fuzzy classification algorithm called fuzzy c-means (FCM) (Bezdek, 1981; Bezdek et al., 1987). The data-clustering problem is to find C clusters in a set of N finite data sets $\{\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^N\}$. The cluster structure can be conveyed by a set of cluster centers $\{\boldsymbol{c}^1, \boldsymbol{c}^2, \ldots, \boldsymbol{c}^C\}$, where $\boldsymbol{c}^i$ is the center of the ith cluster. The minimization of the c-mean objective function can be defined by

$$J = \sum_{k=1}^{N} \sum_{i=1}^{C} (\mu_{ik})^m d_{ik}^2 \tag{4}$$

where d_{ik} is the Euclidean distance expressed by

$$d_{ik}^2 = \|\boldsymbol{x}^k - \boldsymbol{c}^i\|^2.$$

and m is a weight exponent suggested to be within the range 1.5 to 3 by Bezdek (1984). A value of 2 for m is used in this study, and μ_{ik} is a fuzzy membership that measures the degree of association of the kth data point $\boldsymbol{x}^k$ with the ith cluster class:

$$\mu_{ik} = \left(\sum_{j=1}^{C} \left(\frac{d_{ik}}{d_{jk}} \right)^{1/(m-1)} \right)^{-1}, \quad \text{for all } i. \tag{5}$$

Note that

$$\mu_{ik} \in [0,1] \tag{6}$$

$$\sum_{i=1}^{C} \mu_{ik} = 1 \quad \text{and} \quad \sum_{k=1}^{N} \sum_{i=1}^{C} \mu_{ik} = N \tag{7}$$

$$0 < \sum_{k=1}^{N} \mu_{ik} < N. \tag{8}$$

An iterative scheme for solving the optimization problem is given by Bezdek (1981):

$$c^i = \frac{\sum_{k=1}^{N} (\mu_{ik})^m x^k}{\sum_{k=1}^{N} (\mu_{ik})^m}, \quad \text{for all } i. \tag{9}$$

Please note that all variables should be prescaled so that they are put on an equal basis in the analysis in order to avoid the issue of the order-of-magnitude difference between variables. To classify a group of data, the number of classes has to be specified. In general, more classes can help clarify the picture of classification. However, additional classes increase our burden, since we are going to perform test experiments at the clustering centers. The more classes we use, the more experiments we need to perform. Our goal is not to produce the crispiest classification but to locate optima as quickly as possible. Hence an information index is used to determine the optimal number and location of the next set of experiments.

Information Index

Information entropy

According to Shannon's definition (Shannon, 1948; Shannon and Weave, 1949) of information entropy for a variable X, which can randomly take values x from a set X, the information entropy of the set X is

$$S(x) = \sum_{x \in X} p(x) \ln[p(x)], \tag{10}$$

where $p(x)$ is the probability of the event x occurring. If the variable X can only take a narrow range of values, $p(x)$, for these values is close to 1. For other values in X, $p(x)$ is close to 0. Therefore $S(x)$ is close to zero. If the variable X can take a lot of different values in X each time with a small $p(x)$, $S(x)$ will be a large negative number. Therefore, information entropy is a measure of how random a variable is distributed. It decreases when the variable is more randomly distributed.

Let us apply the information entropy to measuring how clearly the ith cluster is classified:

$$S(c^i) = \sum_{x \in X} p(x^k|c^i) \ln[p(x^k|c^i)], \tag{11}$$

where $p(x^k|c^i) = \mu_{ik}/N_i$ is the probability of finding x^k as a representative of the ith cluster and $N_i = \sum_{k=1}^{N} \mu_{ik}$ is defined as the fuzzy number of data of the jth cluster:

$$S(c^i) = \frac{1}{N_i} \sum_{k=1}^{N} \mu_{ik} \ln \mu_{ik} - \ln N_i. \tag{12}$$

Therefore, the entropy of the entire classification set can be defined as

$$S \equiv \sum_{i=1}^{C} \frac{N_i}{N} S_i = \frac{1}{N} \left(\sum_{i=1}^{C} \sum_{k=1}^{N} \mu_{ik} \ln \mu_{ik} - \sum_{i=1}^{C} N_i \ln N_i \right). \tag{13}$$

The second term is a measure of the size of each cluster. If there is one cluster, it is $-\ln N$. If the group is divided evenly into two clusters, then it is $-\ln(N/2)$. The smaller the clusters are, the larger and more orderly the entropy is. Note that $S \to 0$ at $C \to N$. However, the increase in entropy with the increased number of clusters is offset by how clearly we can divide up the data into groups. The term, $\mu_{ik} \ln \mu_{ik}$, represents the penalty of overlapping between groups. If the data points belong to one group (i.e., $\mu_{ik} \to 1$ and $\mu_{jk} \to 0$, $i \neq j$), the contribution of these data points to the first term is negligible. On the other hand, if the data points belong to a large number of groups, μ_{ik} will be nonzero for several groups and the $\mu_{ik} \ln \mu_{ik}$ term becomes significant.

Figure 3a illustrates a group of data points. When the group is classified into different clusters, the size and overlap of the clusters contribute to information entropy (Figure 3b). It can easily be seen that the overlap effect (dark area) is zero when there are exactly four clusters. If we attempt to classify these data into more groups, the informational entropy will actu-

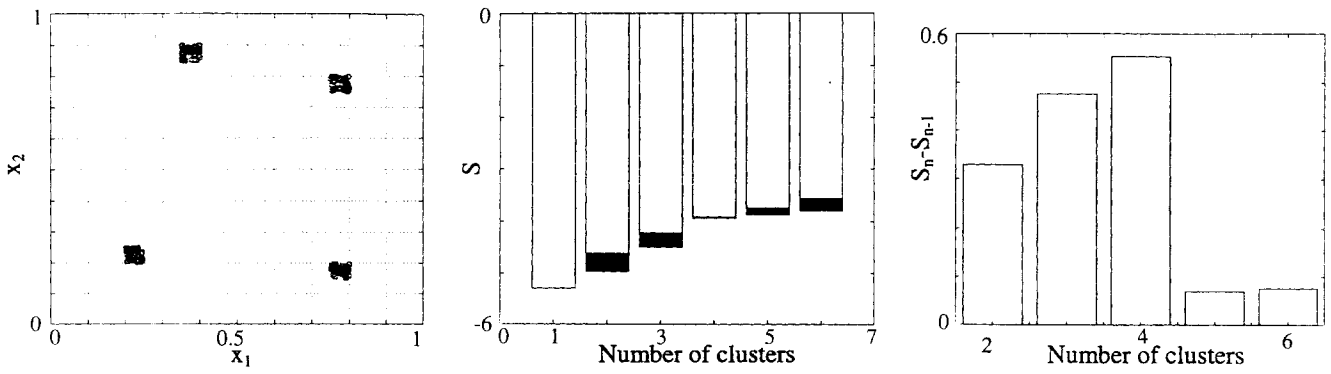

Figure 3. (a) Data to be clustered; (b) information entropy vs. cluster (dark and white parts are the overlap and size contribution, respectively); (c) change in information entropy ($S_n - S_{n-1}$) vs. cluster.

ally increase due to overlap between groups. Although the information entropy keeps increasing with the size of the cluster getting smaller, the increase in information entropy (i.e., gain in knowledge) becomes marginal (Figure 3c), particularly at the change from 4 to 5 clusters.

Information enthalpy

In the previous subsection, information entropy is considered to be a measure for determining a suitable number of clusters. However, the optimization objective is never taken into account. An information energy that is just the expected value of the performance index is defined as

$$U = \sum_{j=1}^{C} \frac{N_j}{N} f[\hat{y}(c^i)] - f_{\min}, \tag{14}$$

where $f_{\min}$ is the value of the minimum f recorded in the optimal search and $f[\hat{y}(c^i)]$ is the performance index evaluated at the cluster centers. The information energy is a measure of the relevance of the messages generated by the clustering analysis to the optimization procedure. Provided we have full confidence in our model, it is most desirable that only one cluster center with objective function close to global is generated. It is also acceptable that the clustering analysis generate a number of centers that have performance indices close to global minima. However, to carry out a lot of experiments at points where the performance expected is poor would be a waste of effort. Please note that the information energy has the unit of the performance index, while information entropy has no unit. The proper scaling of the energy term is discussed in the next subsection. Figure 4 shows a mesh plot of a function and the corresponding contour. The same data points used in the information entropy section are also shown in this contour plot. Figure 5a shows the changes in information energy when we use one to six clusters in the FCM analysis. If we use only one cluster, the center will be located near the point (0.5, 0.5). Due to the large value of the performance index, the information energy is high. While it is expedient from an experimental view to perform only one experiment, information energy tells us that this piece of information is not especially relevant. If we use more clusters, more relevant information will be obtained. If information energy is the only criterion, the number of experiments in the next step should be four because an increase from four to five clusters results in no decrease in information energy, that is, there will be no improvement in the relevance of information obtained.

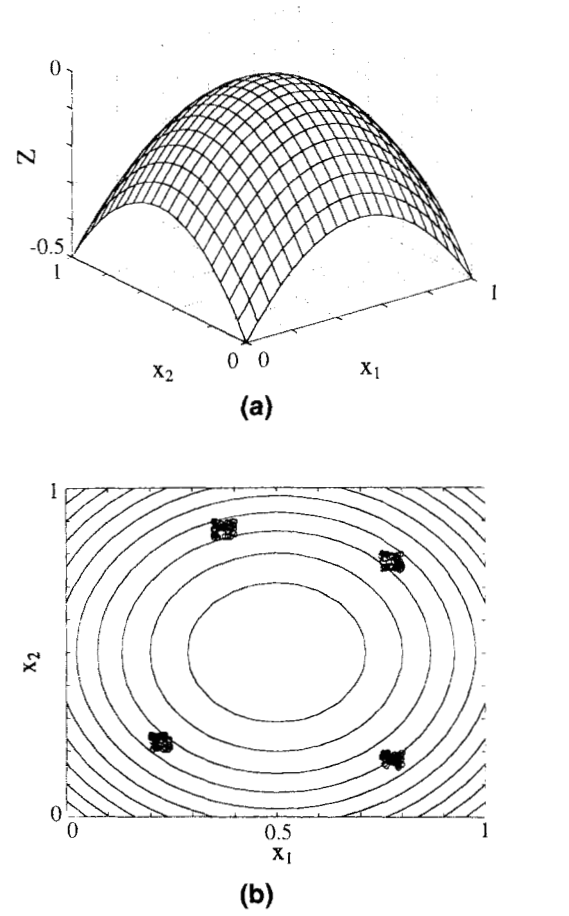

Figure 4. (a) Three-dimensional mesh; (b) group data points against the contour background.

Information free energy

The indices of entropy and energy are measures of how well a set of cluster means classifies the data points and how

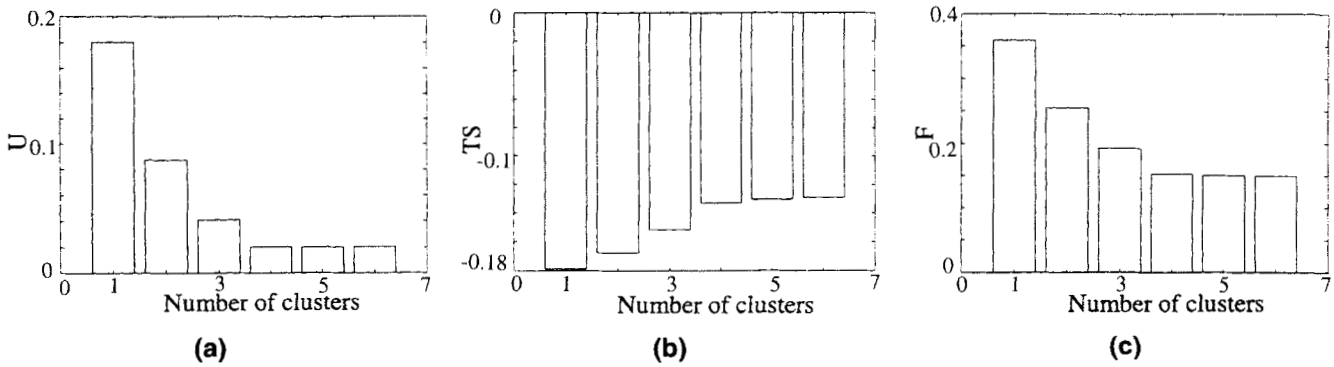

Figure 5. (a) Information enthalpy vs. cluster; (b) enthalpy multiplied by temperature vs. cluster; (c) free energy vs. cluster.

well a set of cluster means performs if it is chosen as the next set of experiments, respectively. However, considering them separately may result in inconsistency between the two. To balance them, a composite information index: the *information free energy* is defined:

$$F = U - TS. \tag{15}$$

The temperature defined is a normalization factor:

$$T = \frac{T_{\max} - f_{\min}}{\ln N}, \tag{16}$$

where $f_{\max}$ is the maximum f of all surviving points in the regional optimal search; and N is the total number of existing experiments. During the procedure of determining the number of clusters, temperature remains constant. This is analogous to the thermodynamic equilibrium criterion under the isothermal condition that the free energy is minimized. During the initial phase of the search, when N is small, the data are relatively scattered, and $f_{\max} - f_{\min}$ is relatively large. We should put more emphasis on obtaining the shape of the performance relation rather than finding the optimum. As the data accumulates with more new experiments, the result of region optimal search will concentrate toward global optima, and $f_{\max} - f_{\min}$ would decrease. Emphasis should be put less on categorizing information and more on optimization.

The changes in information free energy (F) with the number of clusters for the previous 200 data points is shown in Figure 5c. The corresponding two parts [information entropy (S) multiplied by temperature (T), and information energy (U)] are also included. In this case both information energy and information entropy call for four experiments.

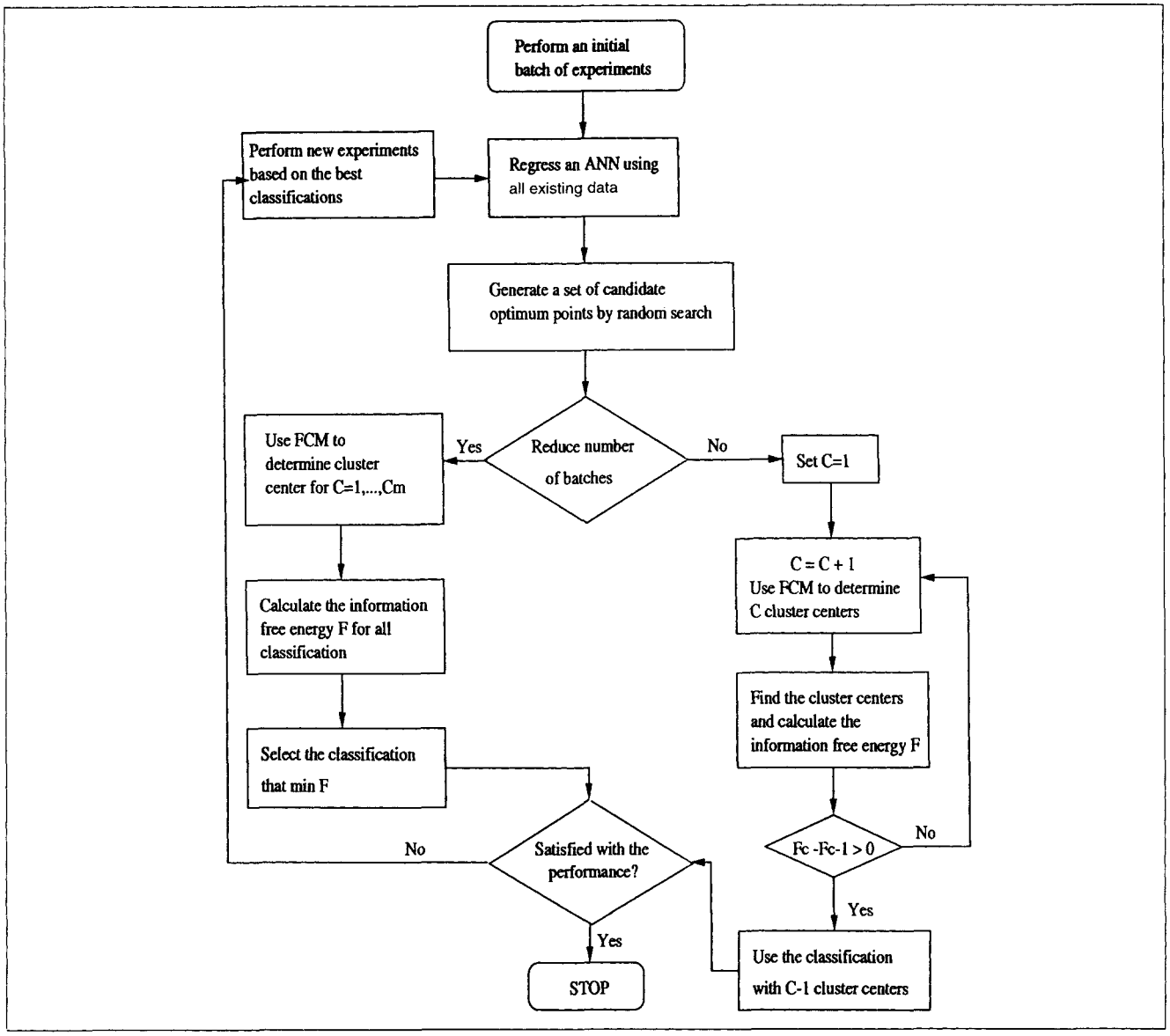

Figure 6. Experimental design.

Implementation of the Information Free Energy Criterion

It is possible for information free energy to show multiple local minima. There are two ways to implement this phenomenon. If the subject of investigation is a recipe of a new product, many tests can be conducted simultaneously in a laboratory environment. The cost of a single experiment is of little concern. We can first decide the maximum number of experiments that can be performed in a single batch, and then calculate the information free energy of each classification. The number of experiments in the next batch should be the one that minimizes information free energy. In the early stages of the search, a relatively large number of experiments will be collected before we update the ANN model, but the total number of batches can be reduced. If the subject of investigation in each single experiment is an expensive step, we can start the classification procedure with just one cluster. Calculate the change of information free energy if another cluster is added. If there is an increase in information free energy, the additional cluster is rejected. Experiments are then performed at the existing cluster centers. If there is a decrease in information free energy, the additional cluster is accepted, and the possibility of adding another cluster is investigated again. When data are scarce, the results of regional optimization will be scattered. The information free energy is likely to decrease when we try to add another cluster. However, the number of experiments in each batch will be kept at the first local minimum of information free energy. The ANN model is updated more frequently. The number of batches may increase, but the total number of experiments will be reduced. A flow chart of the entire experimental design procedure is illustrated in Figure 6.

Case Studies

In order to prove the ability of the proposed method, two examples will be tested. For a simple explanation, a modified Himmelblau function of two independent variables is used because it is easy to visualize a search in a two-dimensional case. The other example is concerned with a practical chemical process of manufacturing ceramic pigment. Both cases to be performed are based on the rule of the reduced number of batches.

Modified Himmelblau function

This function with two independent variables has multiple local optima:

$$Z(x_1, x_2) = (x_1^2 + x_2 - 11)^2 + (x_1 + x_2^2 - 7)^2 + x_1 + 3x_2 + 57 \tag{17}$$

defined for $-5 \le x_1 \le 5$ and $-5 \le x_2 \le 5$. The original Himmelblau function, which consists of only the first two terms of Eq. 17 has four equal local points (Reklaitis et al., 1983). After adding the last three terms of Eq. 17, a global point at $(-3.80, -3.32)$ with a value of 43.3 can be found. From the mesh surface (Figure 7a) and the contour plot (Figure 7b), four local optimal points with a global point are shown. Usually, it is difficult for the traditional experimental design to

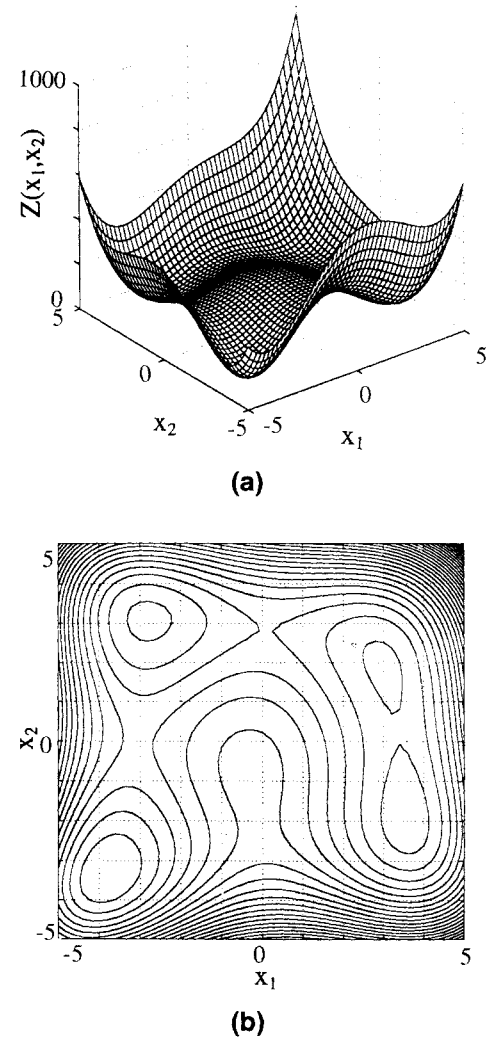

Figure 7. (a) Three-dimensional modified Himmelblau function; (b) the corresponding contour.

build an accurate model for the modified Himmelblau function and to find the global optima. Finding optimal points based on the proposed experimental design method is depicted in Figure 8. A total of 12 batches of experiments are done and the results of the first, the sixth, the ninth, and the twelfth batches are shown. Column (a) lays out the location of the experiments that have been performed against the contour of the Himmelblau function. Column (b) shows the candidates for the next batch of experimental results generated by the regional optimal search against a background of the contour of the current ANN model. Information analysis is given in column (c). The exact locations of the next batch of experiments are shown in column (d).

Note that the information energy is pretty high for all clusters in the first few batches. For example, after the first batch, the current contour shows a single optimum. If decrease in information energy is used as the criterion, only one experiment will be performed. After the sixth batch, two major optima are found and only two experiments would be performed if optimization is our only concern. However, information entropy calls for more experiments that help us to

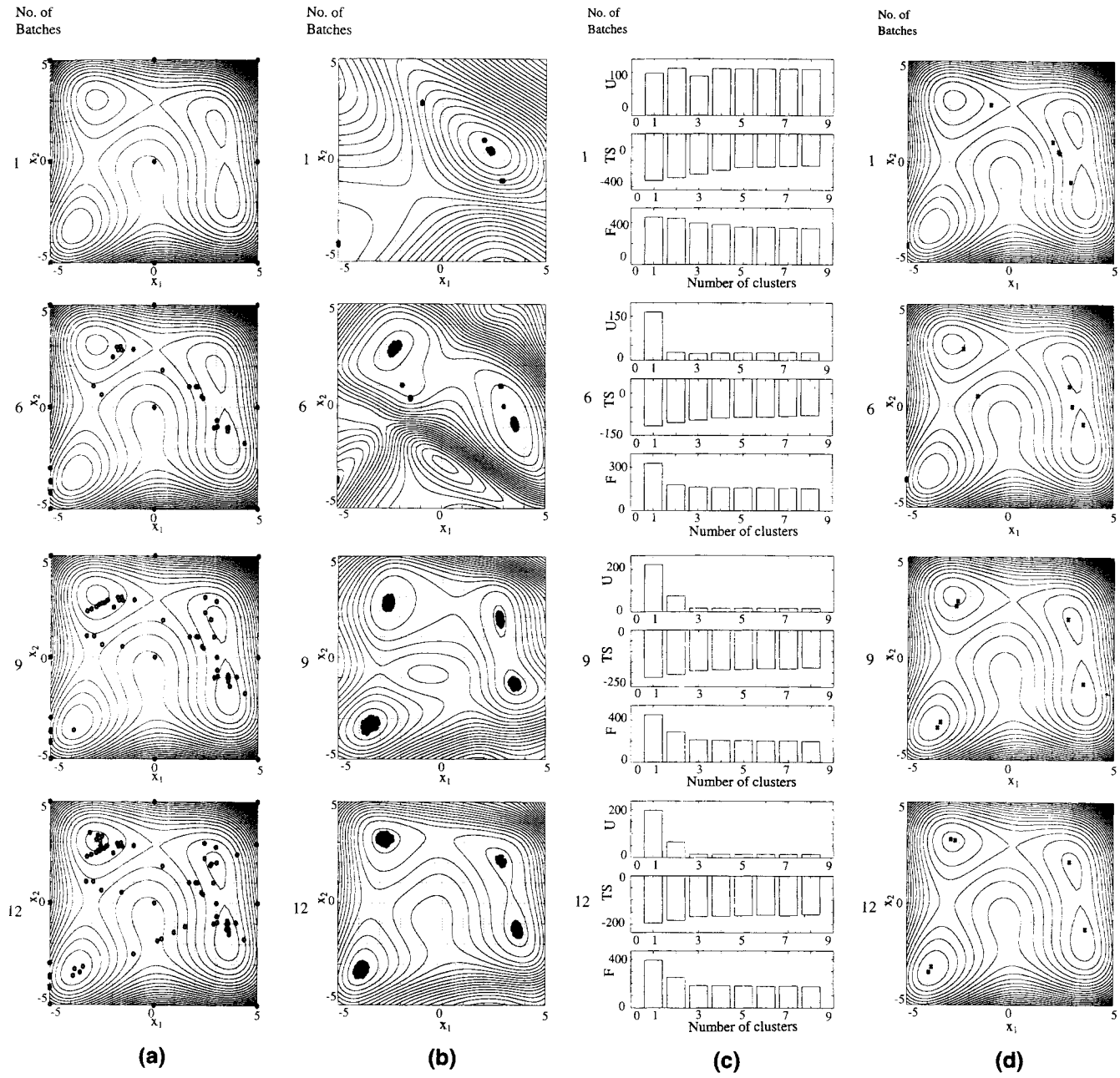

Figure 8. Optimal experimental design at different batches.

(a) The current and the past experimental points against the contour of Himmelblau function. (b) Corresponding model contour whose solid points represent the found local minimum points. (c) Information analysis plot. (d) The next batch of the new experimental points against the contour of Himmelblau function.

mold the performance surface more accurately. In Figure 9, it is found that the representation of the initial ANN model is pretty poor, but the ANN model is approaching the original system little by little with the new experimental data added. In the final stages, the entropic term contributes little to the decision making. Information energy will help us place all experiments around the local optima.

In traditional experimental design, two-level factorial design can only estimate the main effects and interaction and be modeled as a linear plan. In this study, three-level factorial design (3^2) is used to estimate the degree of curvature in the response. The locations of the initial points are shown in Figure 10a. Thus, a second-order response-surface model can be obtained using the least square estimation. The contour of this model is shown in Figure 10b. New experiments are added at the local optima. The response surface is updated to improve the model and generate the new optimal point. However, it is found that the optimal point is trapped into a local point after 40 iterations of searching. Figure 10c displays the result for 40 runs. The failure is due to the inability of a second-order surface to be molded into the proper shape. However, the degrees of freedom of a polynomial response surface increase combinatorially with respect to N. A 4th order 2-variable response surface will have 13 degrees of free-

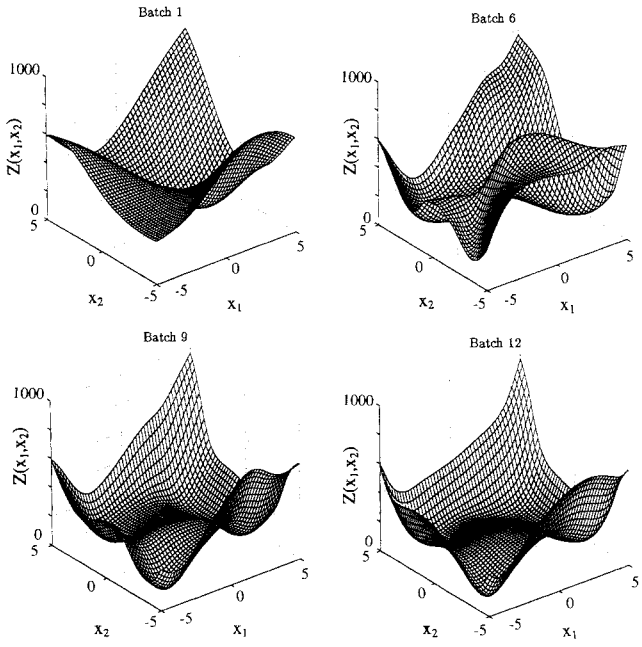

Figure 9. Approximation model at different batches.

dom; hence, a large number of data is needed to obtain a generalizable surface. On the other hand, an ANN can be readily molded into a complex surface, but a feature extraction scheme is needed to screen out the important features that require validation.

Synthesis of cobalt blue color pigment

Aluminum oxide (Al_2O_3), cobaltous oxide (CoO), zinc oxide (ZnO), magnesium oxide (MgO), potassium nitrate (KNO_3), and potassium chloride (KCl) are the basic ingredients of the cobalt blue color pigment. Al_2O_3 is the bulk material of the pigment, and CoO provides the blue color. The color-modifiers, ZnO and MgO, are used during precalcining or premilling to adjust the color of the pigment. The sample can be made greener by adding ZnO, and redder with MgO. Adding mineralizers, KNO_3 and KCl, can reduce the reaction temperature.

The sample preparation and color-measuring process can be outlined as follows:

- The six components are weighted and blended.
- Samples are calcined in a crucible using a preset heating policy. Temperature is ramped to a setpoint and held constant for a long period of time.
- After cooling, the pigments are ground, washed, and dried into particles of small size.
- Body powder and water are added to the pigments. The mixture is powdered in a blender and dried again in an electric-fired kiln.

Three color indices L, a, and b of the sample are measured on a visible spectrophotometer. The color of the final sample is determined by complex interaction between the recipe of the sample and the heating policy in the calcination process. Only the effect of the recipe is presented here. The preceding procedures involve complex chemical reactions. It takes a long time (order of days) to complete the pigment preparation procedure.

The objective is to find a recipe that satisfies the customer's specification:

$$-1.1 \le L = L^{\exp} - L^{\text{ref}} \le -0.9 \tag{18}$$

$$-0.6 \le a = a^{\exp} - a^{\text{ref}} \le -0.4 \tag{19}$$

$$-2.1 \le b = b^{\exp} - b^{\text{ref}} \le -1.9, \tag{20}$$

where the superscripts exp and ref represent the experimental result and the reference point, respectively. The performances of some batch experiments are shown in three-dimensional space (L, a, and b) in column (a) of Figure 11. The results of the suggested experiments are shown in column (b) of Figure 11. The number of experiments and the *best performance* for L, a, and b of each batch are shown in Figure 12. The changes in U, TS, and F with the number of clusters at the first, the sixth, the tenth, and the fourteenth batch of experiment are shown in Figure 13.

The operator's experience provided locations for the first batch of eight experiments, although there might be a large deviation from the desired target. Similarly, in the first few batches, there were suggested experiments that yielded rather

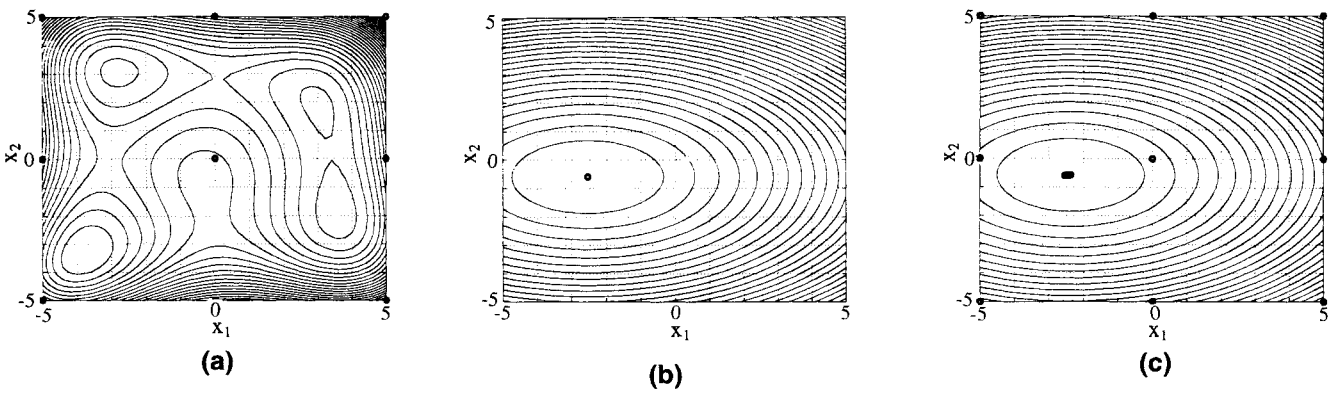

Figure 10. Traditional experimental design using surface response methodology.

(a) Initial experimental points against the counter of Himmelblau function. (b) Optimal point against the counter of the surface response in the first run. (c) All experimental points in 40 batches.

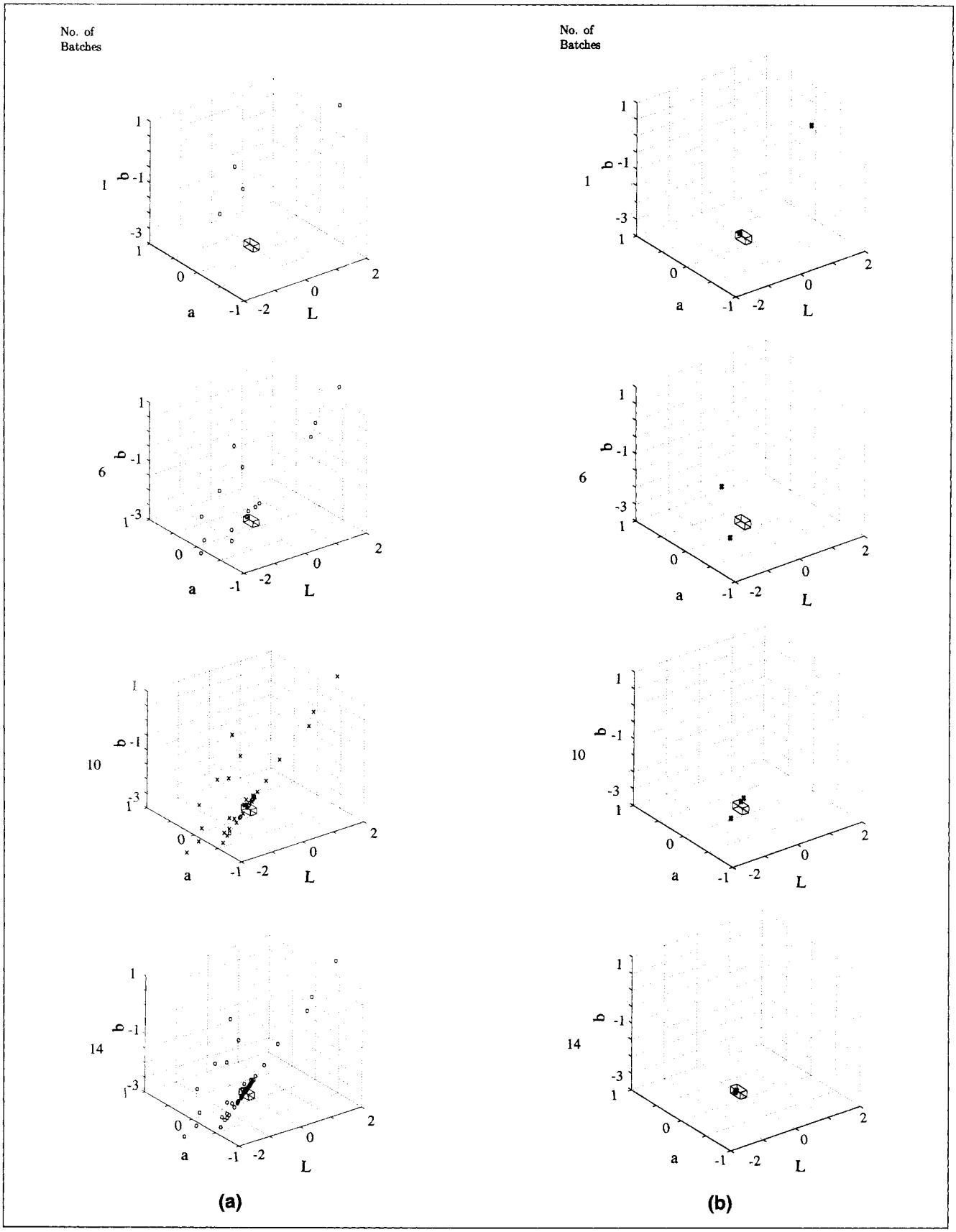

Figure 11. Optimal pigment experimental design at different batches.

(a) Coloring deviation plot for the current and past experiments. (b) Coloring deviation plot for the next experimental points. Note that the cubic box represents the design region, $L = L^{exp} - L^{ref}$, $a = a^{exp} - a^{ref}$, $b = b^{exp} - b^{ref}$.

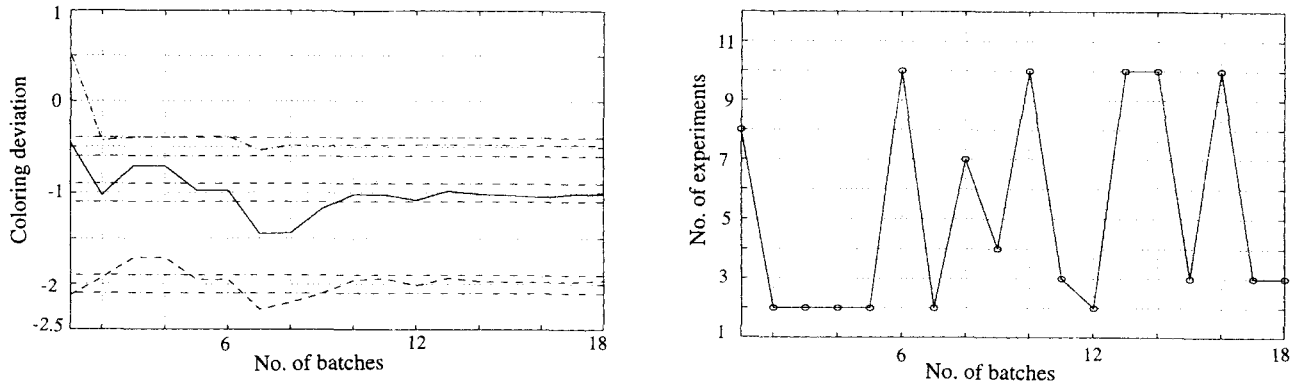

Figure 12. Pigment optimal experimental design path with (a) coloring deviations where *L* = solid line, *a* = dash-dot line, and *b* = dash line; (b) number of experiments needed.

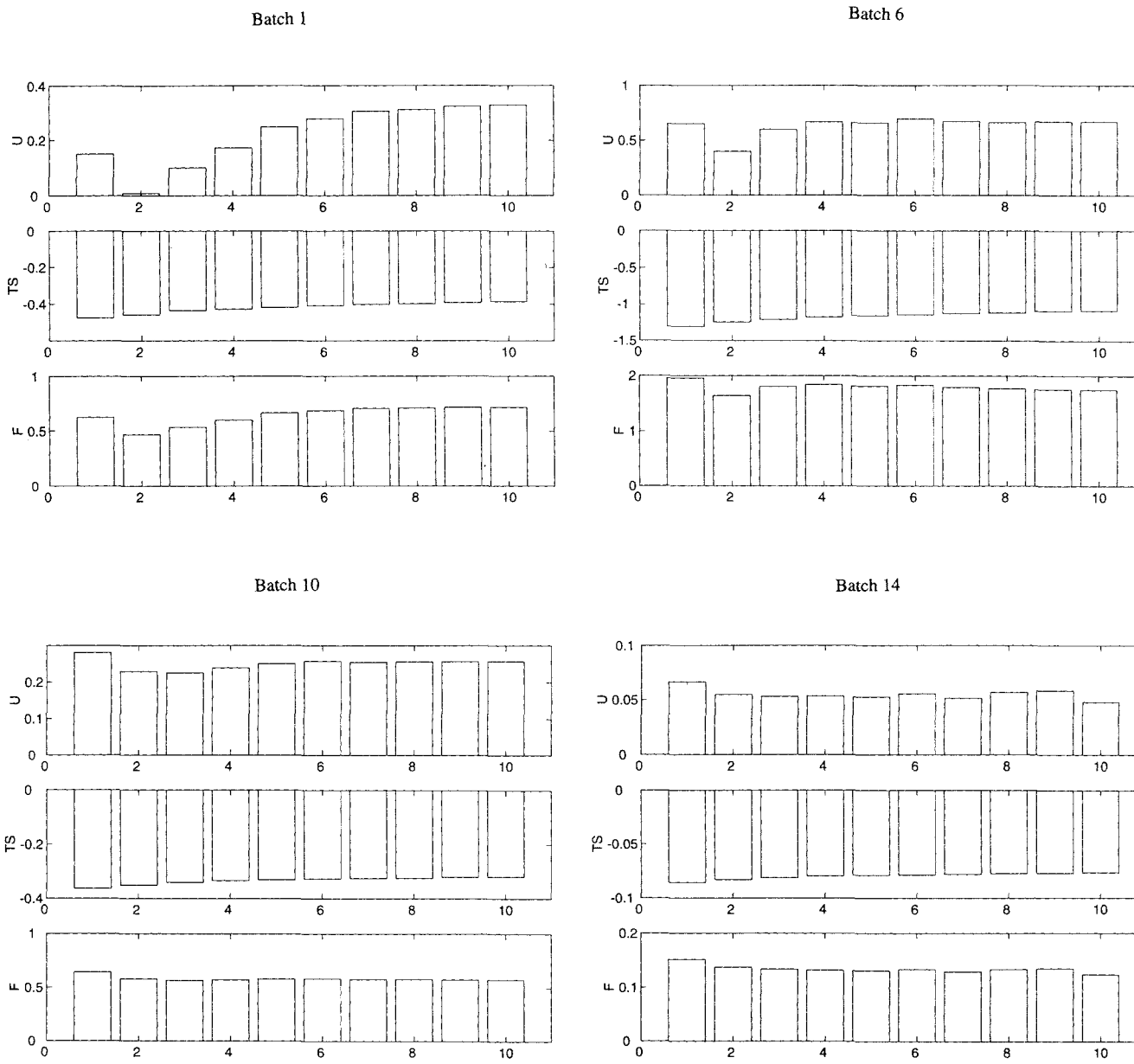

Figure 13. Information analysis plots at different batches.

unsatisfactory results. However, those experimental points were not wasted. They provided information on the response surface that was incorporated into the neural network model. It should be noticed that a feasible recipe was obtained after the fifth batch (Figure 12a). If we allow the procedure to be carried on, it would locate more and more points around the optimal point. After the tenth batch, a near optimum is always found in a new batch. That means more and more points are located around the optimum pigment (Figure 11). The neural network model in this region becomes more and more accurate near the optima. This is particularly important if robustness of the selected operating condition is of great concern. A total of 49 experiments were required to bring us an optimum (after the tenth batch). A full two-level six-factor factorial design required 64 experiments. Since the performance is quite nonlinear, it is unlikely that a 2-level factorial design will locate the correct optimum. A full 3-level 6-factor factorial design required 729 experiments. If a partial factorial design (Myers and Montgomery, 1995) is run using the central composite design, 77 data points were required to perform an initial search.

Conclusion

Quality by design suggests that quality improvement opportunities should be investigated early in the research, development, and design phases. However, the number of variables involved and the complexity of the performance–variable relation of many processes and products limit the applicability of traditional experimental design procedures. A novel experimental design procedure that integrates various elements of artificial intelligence research is proposed. ANN is used as the tool to summarize all experimental information into a mathematical model. Random search is employed to extract features of this model. These features are classified with a fuzzy classification technique and information theory. An information free energy is defined to determine how many features are worth testing. A simulation example demonstrates the logic of the method. An industrial example demonstrates that the proposed method can reduce time, cost, and risk of the product and process development.

Acknowledgment

This work is supported by the National Scientific Council of R.O.C. under Grant NSC85-2622-E007-008 and the pigment blending project sponsored by China Glaze Co., Ltd.

Notation

C = number of clusters
E = performance function for training neural network
h_j = forward signal for passing through the jth hidden neuron
J = objective function for calculating cluster centers
N_j = fuzzy number of data of the jth cluster
P_i = probability of the event occurring
x^p = pth input vector
y^p = pth output vector
y_k = measured output k
$\hat{y}_k$ = predicted output k

Literature Cited

Allen, D. M., "The Relationship Between Variable Selection and Data Augmentation and a Method for Prediction," *Technometrics*, **16**(1), 125 (1974).

Bezdek, J. C., *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum, New York (1981).

Bezdek, J. C., R. Ehrlich, and W. Full, "FCM: The Fuzzy c-Means Clustering Algorithm," *Comput. Geosci.*, **10**, 191 (1984).

Bezdek, J. C., R. J. Hathaway, M. J. Sabin, and W. T. Tucker, "Convergence Theory for Fuzzy c-Means: Counterexamples and Repairs," *IEEE Trans. Syst., Man, Cybern*, **SCM-17**(5), 873 (1987).

Box, G., and N. R. Draper, *Empirical Model-Building and Response Surface*, Wiley, New York (1987).

Fukunaga, K., *Introduction to Statistical Pattern Recognition*, Academic Press, Boston (1990).

Gorodkin, J., L. K. Hansen, A. Krogh, C. Savrer, and O. Winther, "A Quantitative Study Pruning by Optimal Brain Damage," *Int. J. Neural Syst.*, **4**, 159 (1993).

Hertz, J., A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation*, Addison-Wesley, New York (1991).

Hornik, K., M. Stinchcombe, and H. White, "Multilayer Feedforward Neural Networks are Universal Approximators," *Neural Networks*, **2**, 359 (1989).

Hornik, K., M. Stinchcombe, and H. White, "Universal Approximation of an Unknown Mapping and Its Derivatives Using Multilayer Feedforward Networks," *Neural Networks*, **3**, 551 (1990).

Kalman, B. L., and S. C. Kwasny, "Why Tanh? Choosing a Sigmoidal Function," Int. Joint Conf. on Neural Networks, Baltimore, MD (1992).

Lin, J. J., D. S. H. Wong, and S. W. Yu, "Optimal Multiloop Feedback Design Using Simulated Annealing and Neural Network," *AIChE J.*, **41**, 430 (1995).

Lochner, R. H., and J. E. Matar, *Design for Quality, An Introduction to the Best of Taguchi and Western Methods of Statistical Experimental Design*, ASQC Quality Press, Milwaukee, WI (1990).

Myers, R. H., and D. C. Montgomery, *Response Surface Methodology, Process and Product Optimization Using Designed Experiments*, Wiley, New York (1995).

Reklaitis, G. V., A. Ravindran, and K. M. Ragsdell, *Engineering Optimization: Methods and Applications*, Wiley, New York (1983).

Saraiva, P. M., and G. Stephanopoulos, "Continuous Process Improvement Through Inductive and Analogical Learning," *AIChE J.*, **38**(2), 161 (1992).

Shannon, C. E., "A Mathematical Theory of Communication," *Bell Syst. Tech. J.*, **27**, 379 (1948).

Shannon, C. E., and W. Weave, *The Mathematical Theory of Communication*, Univ. of Illinois Press, Urbana (1949).

Taguchi, G., *Introduction to Quality Engineering*, Asian Productivity Organization (1986).

Manuscript received Oct. 22, 1996, and revision received Dec. 23, 1997.