Nonlinear Dynamic Artificial Neural Network Modeling Using an Information Theory Based Experimental Design Approach

Jun-Shien Lin and Shi-Shang Jang*

Chemical Engineering Department, National Tsing-Hua University, Hsin-Chu, Taiwan

In practice, model predictive control is commonly based on a dynamic black-box model. For linear systems, the model is frequently based on either a process system's impulse response or step response. For nonlinear cases, many works have used an artificial neural network (ANN). The quality of the data set used to construct the ANN model is a critical issue. In this work, we present a systematic approach for designing the data set based on information theory. Information entropy is derived to identify the mutual positions among data points in all feasible regions. In addition, information enthalpy is derived to obtain a system's dynamic nonlinearity. Hence, the placements of the new data are designed on the basis of a compromise between the information entropy and the information enthalpy—the information free energy. Also included herein are realistic examples such as pH control. Simulation results demonstrate that the proposed approach is highly promising in terms of obtaining a reliable black-box model, such as ANN, for model predictive control.

1. Introduction

Developing a nonlinear empirical black-box model for model predictive control (MPC) has received extensive interest, with notable examples including an artificial neural network (ANN)[1,2] or a rule-based model.[3,4] However, to our knowledge, the data that comprise the black-box model have never been investigated. Therefore, in this work, we develop a systematic approach, i.e., the information free energy design (IFED), to experimentally design a dynamic system. The proposed IFED considers the distribution of the data in the feasible region and the topology of the dynamic system such that the black-box model can accurately predict the system's dynamic behavior well.

Linear model predictive control has been successfully employed in process industries in recent decades. A "good" model must be obtained through plant tests to implement this control scheme. Step response models[5] and impulse response models[6] have been implemented. In this area, all applicable approaches are based on nonphysical meaning models or so-called black-box models. Garica and Morari[7] thoroughly reviewed the linear model predictive theory, including parametric and nonparametric models. Only nonparametric models, e.g., dynamic matrix control (DMC), are implemented in most industrial processes.

Theoretical development of nonlinear model based control has not yet been mature. Jang et al.[8] extended linear MPC to the nonlinear area using physical models. Their investigation assumed that a physical model, with some unknown parameters, can be obtained through material and energy balances. In addition, the unknown parameters can be identified on-line. However, in most cases, obtaining the physical meaningful model is relatively difficult. The feasibility of applying black-box models such as ANN has received increasing attention. In process control, implementing such a model for model predictive control is common practice. Our previous works[3,4] directly implemented the dynamic data set of processes as the process model. In addition, numerous attempts have been made to implement back-propagation networks[9] or recurrent neural networks[10] for nonlinear MPC. Most works implement a one-step-ahead ANN model in the area of MPC, but some works implement a multistep ANN model rather than the one-step-ahead model.[1,11] However, of paramount concern is that, for linear MPC, many systematic experimental approaches ensure the qualities of the black-box model, such as DMC. However, no such technique is available for nonlinear systems. Therefore, in this work, we present a systematic approach to ensure the quality of the ANN model.

Experimental design for general systems is essential for product developments. Statistical approaches, with strong theoretical backgrounds, are frequently implemented for such developments. However, for nonlinear systems, many useful statistical theories cannot be implemented. Our recent work[12] applied ANN toward experimental design. However, that work focused only on obtaining high-quality products using the least possible number of experiments. In this work, we construct a reliable model for model predictive control. In this case, reviewing the complete topology of the plant dynamics is highly desired.

In developing a trustworthy plant dynamic ANN model, this work assumes the notion that a temporary ANN can be constructed on the basis of the available experimental data set. Newly designed experimental data can be developed using that temporary ANN. Reviewing the complete plant dynamic topology involves attaining the largest possible feasible boundaries of the process dynamics based on the limited available current data. Substantial development has recently been made in Delaunay triangulation, which was originally developed in the 1970s.[13] The Delaunay triangulation algorithm triangulates the arbitrary plane domain. A convex hull of existing points can be recognized through the triangulation. This approach has recently been

* To whom correspondence should be addressed. Telephone: 886-3-5713697. Fax: 886-3-5715408. E-mail: ssjang@che.nthu.edu.tw.

S0888-5885(98)00124-9 CCC: $15.00
Published on Web 07/22/1998

extended to 3-D triangulation[14] and other nonconvex pattern representations.[15]

Shannon[16] derived information entropy to measure the uncertainty of a random variable. Hence, evaluating the quality of an ANN model hinges on implementing the information entropy.[17] In this work, we implement the information entropy to measure the uncertainty of a possible event in the black-box model. Also derived herein is the information enthalpy to measure the "nonlinearity" of a possible event. The optimal placements of the experiments therefore depend on the compromise of the above entropy and enthalpy—information free energy.

The pH control system has been widely discussed among chemical process control researchers due to its severe nonlinearity. In most cases, PID controllers do not function properly in these systems. Many works concentrated on model predictive control of pH plants using a complete first principle dynamic model[18] or partial plant knowledge adaptive control.[19] Proll and Karim[20] implemented a nonphysical meaning polynomial model to perform model predictive control of a pH system. Pottmann and Seborg[21] implemented an ANN model to perform on-line model predictive control of a pH plant. In this work, we select a detailed, simulated pH plant, derived and tested by Hensen and Seborg,[18] as an example. According to our results, IFED is highly promising for developing a dynamic ANN model for the purpose of nonlinear model predictive control for this pH process.

The rest of this paper is organized as follows. Section 2 presents the problem statement and formulation. Section 3 provides the development of information free energy. Section 4 presents the algorithm of the information free energy experimental design and the Delaunay triangulation of the feasible region. Numerical examples, including a realistic pH control system, are illustrated in section 5. Concluding remarks are finally made in section 6.

2. Problem Statement

A dynamic model predicts the plant output in the future using the past trajectories of input and output. Assume that a single-input−single-output (SISO) dynamic system can be discretized as follows:

$$y_{k+1} = f(y_k, y_{k-1}, ..., y_{k-n}, u_k, ..., u_{k-m}) \tag{1}$$

where n and m denote the output and input orders of a dynamic system, respectively. A standard model predictive control problem, given a dynamic model (1), can be formulated as the following:

$$\underset{u_{k+1}, u_{k+2}, ..., u_{k+p}}{\text{Min}} \sum_{j=1}^{p} \varphi_j (\hat{y}_{k+j} - y_{d,k+j})^2 + \gamma_j \Delta u_{k+j}^2$$

s.t. (1)

$$u_{\min} \le u_k \le u_{\max}$$

$$|\Delta u| \le \Delta u_{\max} \tag{2}$$

where $\hat{y}_{k+j} = y_{k+j} - \tilde{y}_{k+j}$, y_{k+j} is the on-line measurement, $\tilde{y}_{k+j}$ is the model output, $y_{d,k+j}$ is desired system output, $p \le q$, such that $u_{k+p} = u_{k+p+1} = ... = u_{k+q-1}$, and φ_j and γ_j are weightings of the objective function. The numerical solution of (2) is easy, as discussed in our previous paper.[8] However, the key issue is the development of a reliable dynamic model (1). This work is aimed at obtaining an ANN model based on a systematic approach. A multi-input−multi-output (MIMO) model predictive control is a much more complicated control problem than a single-input−single-output (SISO) process system. In this work, only SISO examples are discussed. Notably, the approach derived here is also useful for determining a MIMO model.

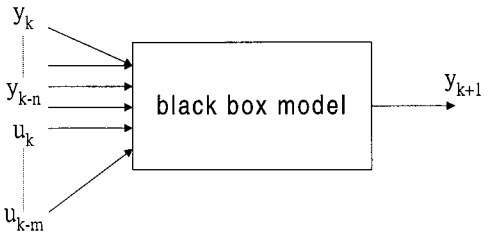

Figure 1. Inputs and output of an ANN model presenting the system dynamics of a SISO plant.

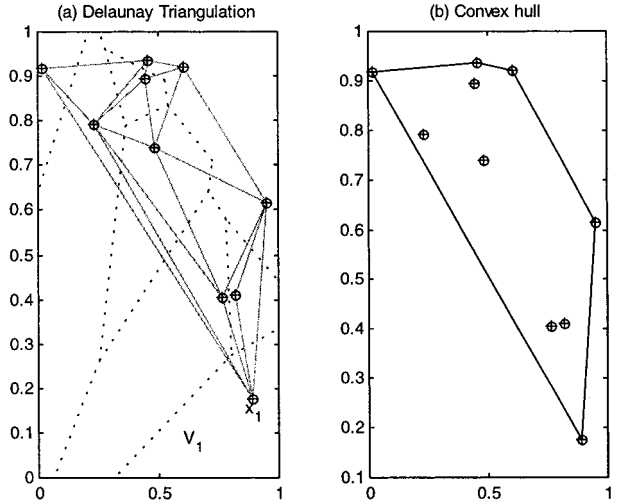

Figure 2. Delaunay triangulation and convec hull: (a) - - -, Voronoi polygon; —, Delaunay triangulation; circled plus, given data points; (b) —, convex hull; circled plus, given data points.

Define θ

$$\theta = (u_k, y_{k-1}, ..., y_{k-n}, u_k, ..., u_{k-m}) \tag{3}$$

as an event of the dynamic space. Also, since y_{k+1} is uniquely determined by (1), we denote the following corresponding augmented event:

$$\mathbf{x} = (y_{k+1}, y_k, y_{k-1}, ..., y_{k-n}, u_k, ..., u_{k-m})$$

Notably, determining n and m is a nontrivial task. This issue has been addressed in our previous work[3] as well as others.[22] For simplicity, we assume that n and m in (1) have already been determined.

The experimental design attempts to determine function f in (1), implicitly by a black-box model, e.g., ANN. Figure 1 depicts the inputs and output of the black-box model. Given a set of ***testing data***

$$\Psi = \{x_i = (y^i_{k+1}, y^i_k, ..., y^i_{k-n}, u^i_k, ..., u^i_{k-m})|_{i=1,...,M}\} \tag{4}$$

Herein, we attempt to find a set of ***training data*** for the black-box model

$$\Omega = \{x_i = (y^i_{k+1}, y^i_k, ..., y^i_{k-n}, u^i_k, ..., u^i_{k-m})|_{i=1,...,N}\} \tag{5}$$

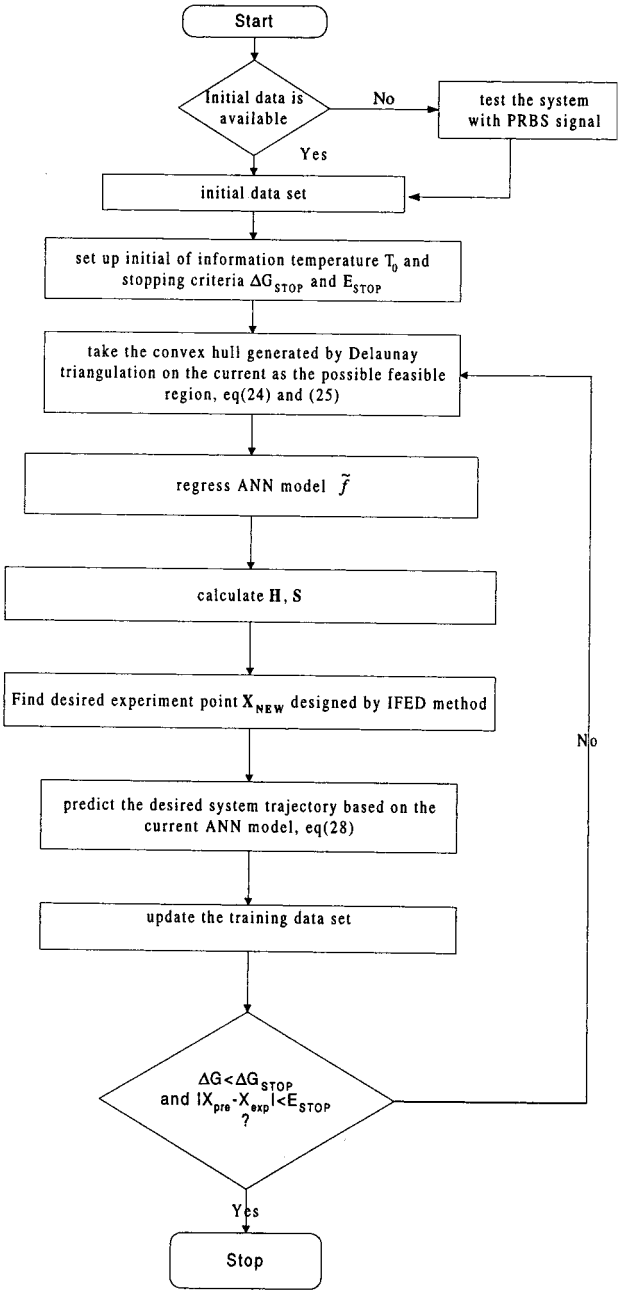

Figure 3. Computer flowchart of IFED.

such that function f in (1) can be determined and the following objective function can be minimized:

$$\sum_{i=1}^{M}(y_{k+1}^i - \tilde{y}_{k+1}^i)^2$$

$$\text{s.t.} \qquad \tilde{y}_{k+1}^i = f(y_k^i, y_{k-1}^i, ..., y_{k-n}^i, u_k^i, ..., u_{k-m}^i) \tag{6}$$

where $i = 1, ..., M$. That is, the prediction error of the process output can be minimized in the testing set. In here, the following constraints are implemented due to the limitations of the control resources.

$$u_{\min} \le u_{k-i} \le u_{\max} \tag{7}$$

$$|\Delta u_{k-i}| \le \Delta u_{\max} \tag{8}$$

Since the control actions are subjected to (7) and (8), the output states cannot go anywhere. The output must

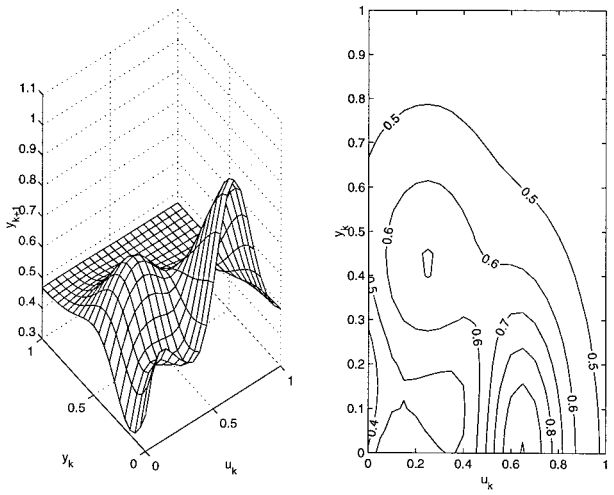

Figure 4. 3-D plot and the contour of the dynamic model of example 1.

be bounded, and these bounds are functions of the previous state. Therefore, designing the training set is also subjected to the following constraints:

$$y_{k-j,\max} = h(y_{k-j-1}) \tag{9}$$

$$y_{k-j,\min} = h'(y_{k-j-1}) \tag{10}$$

where $i = 0, 1, 2, ..., m - 1$, $j = 1, 2, ..., n - 1$. Herein, we denote the following feasible event set:

$$\Phi = \{x | x = (y_{k+1}, y_k, ..., y_{k-n}, u_k, ..., u_{k-m}), x \text{ satisfies (1), (7), (8), (9), and (10)}\} \tag{11}$$

However, the feasible event set is continuous and impossible to obtain numerically. In this work, Φ can be more feasibly approximated by discretizing the possible event set as follows:

$$\Phi' = \{x_{I_0,I_1,...,I_n,J_0,...,J_m} | x_{I_0,I_0,...,I_n,J_0,...,J_m} = (y_{k+1}, y_k^{I_0}, ..., y_k^{I_n}, u_k^{J_0}, ..., u_k^{J_m}), x_{I_0,I_1,...,I_n,J_0,...,J_m} \in \Phi\} \tag{12}$$

where $1 \le I_0, I_1, ..., I_n, J_0, J_1, ..., J_m \le Q$. The total amount of the approximate feasible events can be determined:

$$P = Q^{n+m}$$

The optimization problem cannot be directly solved unless the dynamic model (1) is attained. The following sections provide another means of solving this experimental design problem. It should be noted that the determination of a feasible set depends on the determination of functions h and h' in (9) and (10). However, h and h' cannot be explicitly obtained due to the lack of plant knowledge. In this work, h and h' are implicitly found by implementing the existing data set Ω and the approach of Delaunay triangulation as described in section 4.1.

3. Development of Information Free Energy Design (IFED)

3.1. Information Entropy. Shannon[16] first derived information entropy. The information entropy of a

discrete random variable Z taking value of z is defined by

$$S(Z) = \sum_z p(z) \ln[p(z)] \tag{13}$$

where $p(z)$ denotes the probability density of the event $Z = z$. If Z only takes narrow values, then $p(z)$ is close to 1; for other values of Z, $p(z)$ is close to zero. In this case, $S(z)$ is close to zero. If Z can take many different values in each time with a small $p(z)$, then $S(z)$ becomes a large negative value. In the case of a continuous random variable Z, the information entropy for this event can be determined:

$$S(Z) = \int_{-\infty}^{\infty} p(z) \ln[p(z)]\, dz \tag{14}$$

Given existing experimental data $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_i, ..., \mathbf{x}_N \in \Omega$, we denote the following probability measure of uncertainty of an event $\mathbf{x} \in \Phi$:

$$p(u) = \frac{1}{\sqrt{2\pi}\sigma} e^{-|u-f(\theta)|^2/2\sigma^2} \tag{15}$$

where u denotes the model prediction of $f(\theta)$ and takes the value from $-\infty$ to ∞ and

$$\sigma^2 = \frac{1}{\sum_{i=1}^{N} \mu(x|x_i)} \tag{16}$$

denotes the standard deviation of the Gaussian distribution of the random variable $\mathbf{x}$, where $\mu(x|x_i)$ is a fuzzy membership measure of event $\mathbf{x}$ belonging to experiments $\mathbf{x}_i$. Here, we assume that $\mu(x|x_i) = \mu(d = |x - x_i|)$; i.e., the membership function is only a function of distance between the possible event and the experiments. A nearer event to an experiment takes a larger membership, as suggested in the appendix. Geometrically, the probability distribution of a possible event $\mathbf{x} \in \Phi$ becomes "sharper" if more experiments performed are nearby. Correspondingly, the information entropy for a possible event can be determined as follows:

$$S(\mathbf{x}) = \int_{-\infty}^{\infty} p(u) \ln p(u)\, du = -\frac{\ln(2\pi\sigma^2)}{2} \tag{17}$$

Now, the total information entropy for all events in the approximate possible event set Φ' can be determined:

$$S = \sum_{i=1}^{P} S(\mathbf{x}_i) \tag{18}$$

According to the definition of information entropy, the events in Φ' are more "certain" if the total information entropy is maximized by the design of the training set Ω. However, the information entropy only analyzes the positions of the training data set Ω; i.e., the maximization of information entropy attempts to place the experimental points in proximity to where the training data are "sparse".

3.2. Information Enthalpy. In this work, we denote the "gain variation" of the temporary ANN model as the information enthalpy measure for an event $\mathbf{x} \in \Phi$. The nomenclature "enthalpy" is adopted herein to correlate with the nomenclature "information entropy"

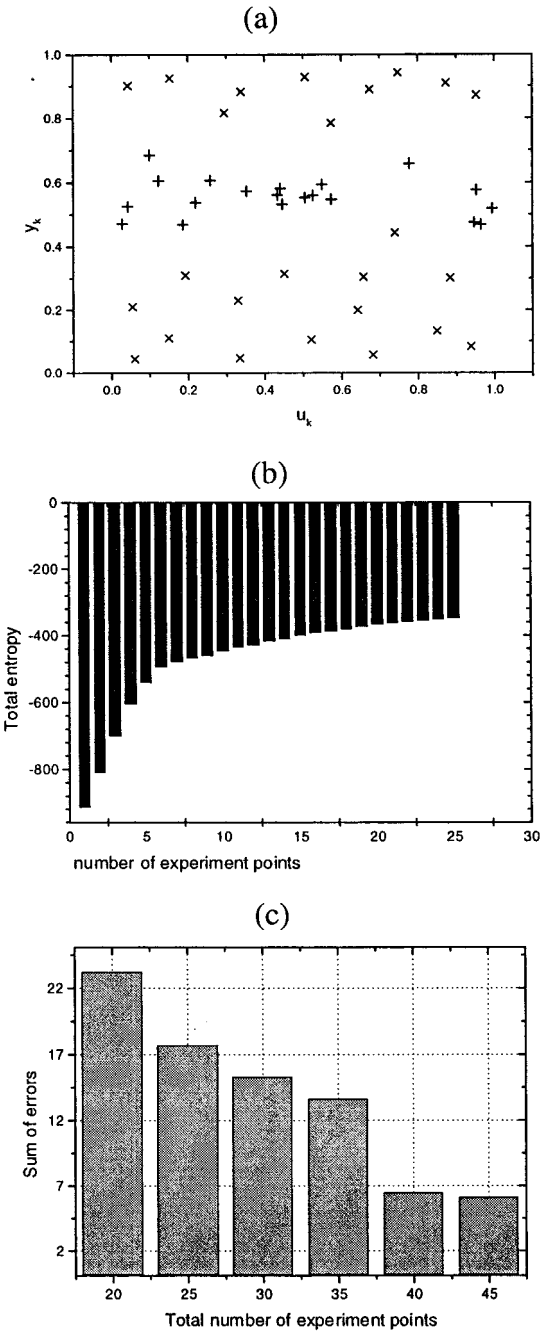

Figure 5. Entropy-based experimental design of example 1. (a) The placements (×) of the experimental data. (b) The total entropy of the elements of Φ' vs number of experiments. (c) The total prediction error of the ANN model to the elements of Φ' vs total number of experiments, where the symbols (+) are the initial data.

derived in a previous investigation.[16] Assume that an implicit relation $\tilde{f}$ between the input and output sets can be determined using the existing training data set:

$$y_{k+1} = \tilde{f}(y_k, y_{k-1}, ..., y_{k-n}, u_k, ..., u_{k-m}) \tag{19}$$

Then, for all events in the feasible event set, $\mathbf{x} \in \Phi$, the following information enthalpy measure is defined:

$$H(\mathbf{x}) = \text{norm}\left(\frac{\partial^2\tilde{f}}{\partial y_k^2}, \frac{\partial^2\tilde{f}}{\partial y_{k-1}^2}, ..., \frac{\partial^2\tilde{f}}{\partial y_{k-n}^2}, \frac{\partial^2\tilde{f}}{\partial u_k^2}, ..., \frac{\partial^2\tilde{f}}{\partial y_{k-m}^2}\right) \left(1 + \frac{1}{\sigma^2}\right) \tag{20}$$

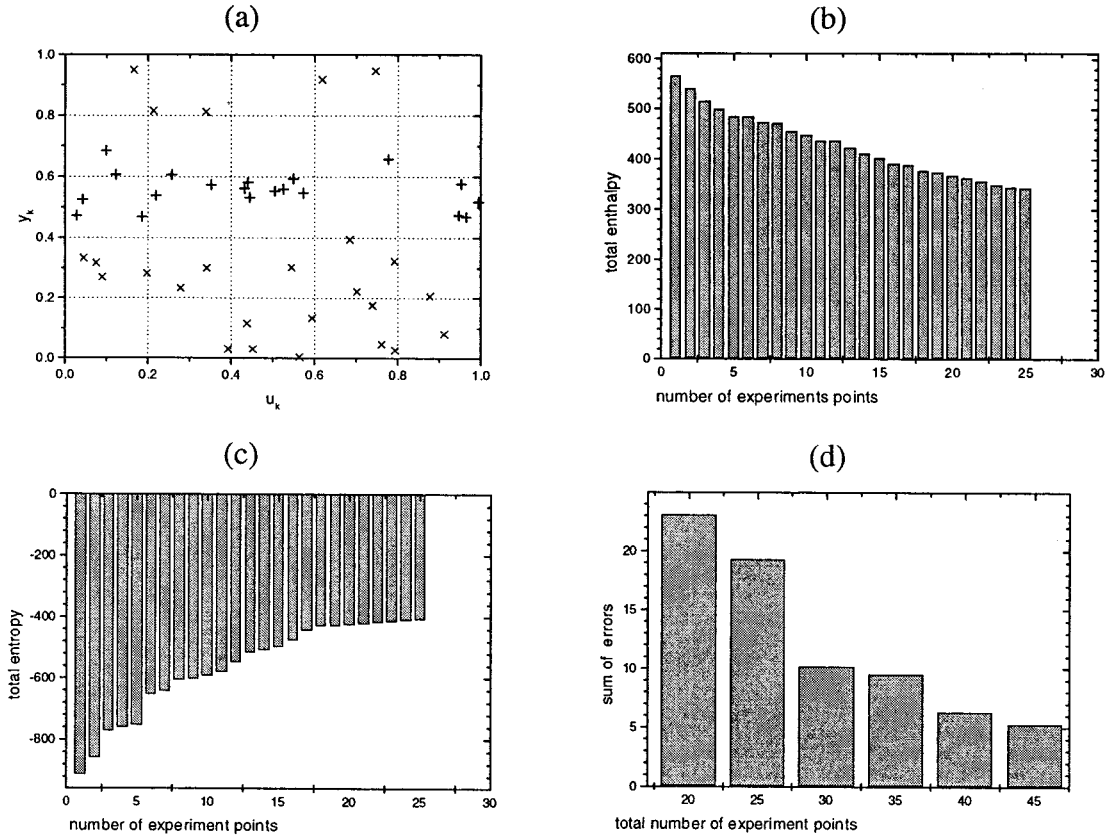

Figure 6. Enthalpy-based experimental design of example 1. (a) The placements (×) of the experimental data. (b) The total enthalpy of the elements of Φ' vs number of experiments. (c) The total entropy of the elements of Φ' vs total number of experiments, where the symbols (+) are the initial data.

where σ^2 is denoted as in eq 16. Then, for all points in $\mathbf{x}_i \in \Phi'$,

$$H = \sum_{i=1}^{P} H(\mathbf{x}_i) \tag{21}$$

According to (20), geometrically, the event that the change in the gradients of outputs to inputs is larger implies that the information enthalpy measure for the event is also larger. Therefore, we perform experiments in these places since a black-box model is nothing but a "nonlinear approximation" to a real plant model (1). An experiment on a "nonlinear" place will also efficiently decrease the total information enthalpy (21). Therefore, the objective of minimization of (21) is equivalent to finding the "nonlinear" place of the approximate model $\tilde{f}$ if there is no point at that place. The term $1 + 1/\sigma^2$ in (20) is to include the effect of existing nearby experiments.

3.3. Information Free Energy. The placement of a newly designed experimental data attempts to make all possible events more "informative". However, if the plant nonlinearity is significant, more experiments should be performed in such areas. Therefore, we not only want to maximize the information entropy but also want to minimize the information enthalpy. The following free energy should hence be minimized:

$$G = H - TS \tag{22}$$

where T denotes the information temperature. The experimental data set is initially assumed to be rather "sparse". Plant dynamics cannot be accurately predicted by (19) at that time. The information temperature should be extremely high and, therefore, we only implement the information entropy rather than information enthalpy. As more experimental data are accumulated, the information temperature should be lower since the plant model becomes much more accurate. This phenomenon closely resembles the concept of simulated annealing development in thermodynamics.[23] The following definition of information temperature is hence derived:

$$T(N) = T_0 \exp(-cN^{1/z}) \tag{23}$$

where N denotes the total number of experiments, as defined in (5). The initial temperature, T_0, reflects how sparse the initial training data set is. If the initial data are not uniformly spread in the feasible region, we recommend that T_0 should be large. Otherwise, if the initial data set is already uniformly spread in the feasible region, then the initial temperature can be small. In other words, the information free energy design implements the information entropy initially if the initial training data set is sparse. In addition, c and z are tuning factors. Further details can be found in Jang et al.[23] In this work, c and z are set to be constants in all three examples in section 5.

4. Information Free Energy Experimental Design for Dynamic Systems

4.1. Initial Model and Feasible Region Buildup. Consider a dynamic system (1), in which it is reasonable

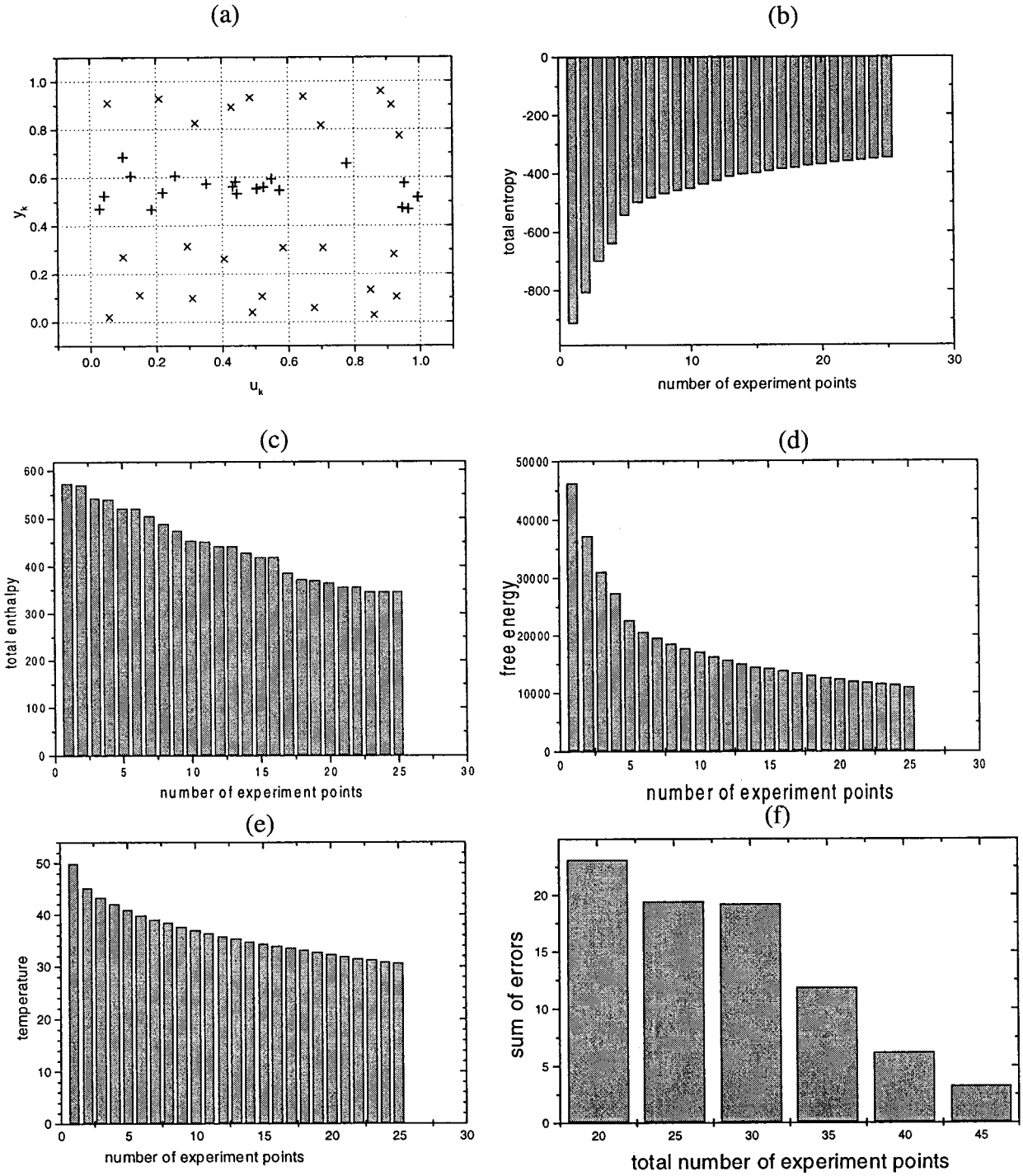

Figure 7. IFED experimental design of example 1. (a) The placements (×) of the experimental data. (b) The total entropy of the elements of Φ′ vs number of experiments. (c) The total enthalpy of the elements of Φ′ vs total number of experiments. (d) The total information free energy of the elements of Φ′ vs number of experiments. (e) Information temperature vs number of experiments. (f) The total prediction error of the ANN model to the elements of Φ′ vs total number of experiments, where the symbols (+) are the initial data.

to assume that the training data set Ω_0 is initially a nonempty set if the plant has been operated for a while. If Ω_0 is empty, then a random input sequence such as a pseudorandom binary sequence (PRBS) can be generated to obtain a "fair" initial training data set Ω_0. However, the PRBS training data set is good only for a linear system.[24] In nonlinear systems, the initial data set generated by PRBS cannot be used for constructing a black-box model that can be used to predict the output y_{k+1} by (1). The feasible region can be roughly obtained by the following approach.

Current state y_k is bounded by the previous state y_{k-1} due to the limited control resource:

$$y_{k,\max} = h(y_{k-1}) \tag{24}$$

$$y_{k,\min} = h'(y_{k-1}) \tag{25}$$

The above functions h and h' cannot be obtained explicitly due to the lack of plant knowledge. However, these functions can be found by the following Delaunay triangulation.

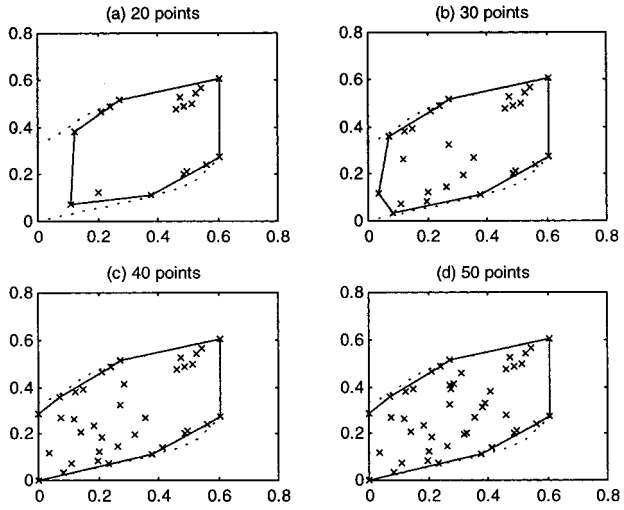

Figure 8. Attainable dynamic feasible region (solid linea) of example 2 using the Delaunay triangulation based on (a) 20 IFED data points, (b) 30 IFED data points, (c) 40 IFED data points, (d) 50 IFED data points, compared with the analytical feasible region (dashed lines).

In light of the initial data set Ω_0, the Delaunay triangulation[25] proposes that all points $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k, ..., \mathbf{x}_N\} = \Omega_0$, denote corresponding Voronoi polygons $\{V_1, ..., V_k, ..., V_N\}$ in such a way that V_k is a territory that is closer to $\mathbf{x}_k$ than any other points in Ω_0; therefore:

$$V_k = \{p \in V_k: \ ||p - x_i|| < ||p - x_j||, \quad \forall j \neq i\} \tag{26}$$

The set of all Voroni polygons is called a Voronoi diagram, as shown in Figure 2a. The connection of data points, whose Voronoi polygons are joined, results in a triangulation, as shown in Figure 2a. The region of a convex hull by conciliation of all jointed edges of two triangles is the possible feasible region in Figure 2b. For details, the readers are referred to MathWorks.[25]

4.2. Placements of Experimental Data. The placements of the experimental data are to minimize the information free energy (22). Assume that an initial data set Ω_0 and a feasible region set Φ' exist. Let us denote a new data $\mathbf{x}_{\text{NEW}}$, which is to be added into the data set Ω_0, such that $\Omega = \Omega_0 \cup \{\mathbf{x}_{\text{NEW}}\}$, and let G be information free energy for new training set Ω and the approximate feasible region set Φ', denoted by (22). The following problem should be solved for the placement of $\mathbf{x}_{\text{NEW}}$:

$$\underset{\mathbf{x}_{\text{NEW}}}{\text{Min}}\ G$$

$$\text{s.t.} \quad \bar{y}_{k+1} = \tilde{f}(y_k, y_{k-1}, ..., y_{k-n}, u_k, ..., u_{k-m})$$

$$\bar{y}_{k-j,\max} = h(\bar{y}_{k-j-1})$$

$$\bar{y}_{k-j,\min} = h'(\bar{y}_{k-j-1})$$

$$u_{\min} \le u_{k-1} \le u_{\max}$$

$$|\Delta u_{k-i}| \le \Delta u_{\max} \tag{27}$$

where $\mathbf{x}_{\text{NEW}} = (\bar{y}_{k+1}, \bar{y}_k, \bar{y}_{k-1}, ..., \bar{y}_{k-n}, u_k, ..., u_{k-m})$ and $\tilde{f}$ is a temporary ANN model derived (trained) using the existing Ω_0. Notably, the entire trajectory $\bar{y} = (\bar{y}_{k+1}, \bar{y}_k, ..., \bar{y}_{k-n+1}, \bar{y}_{k-n})$, must be designed for $\mathbf{x}_{\text{NEW}}$. Hence, the dynamic system must be driven to follow the desired trajectory using the ANN model derived from Ω_0.

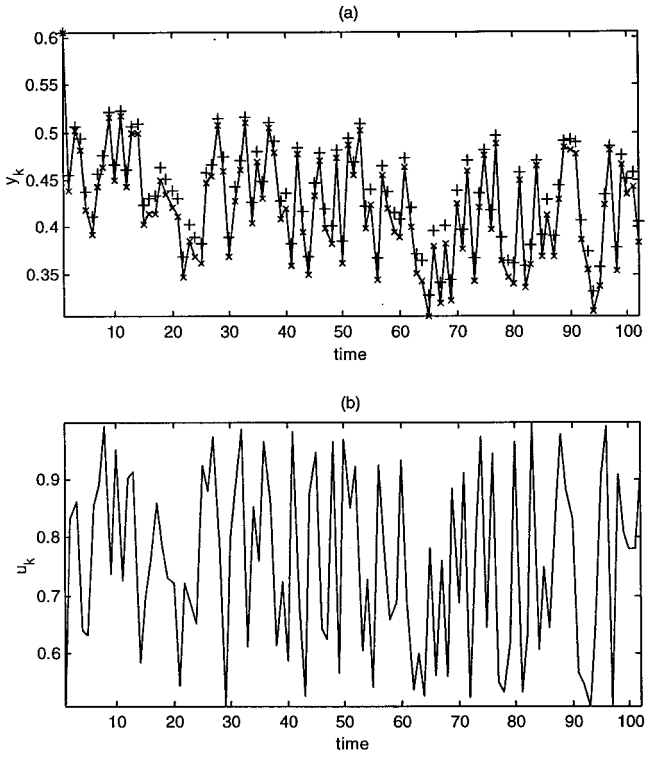

Figure 9. Trajectory following of the ANN model: ×, ANN trained by 50 IFED data; —, system output; +, ANN trained by 20 IFED points, to the plant output.

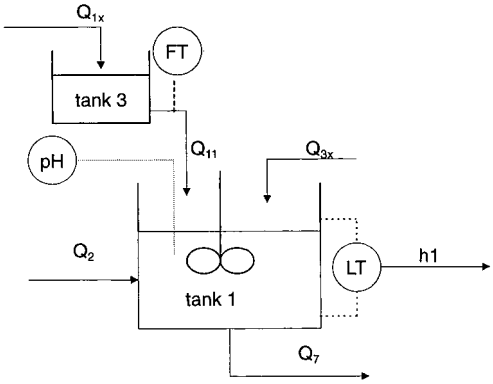

Figure 10. Schematic plot of the pH control system.

In most cases, in dynamic system (1) the output order is greater than the input order, i.e., $n \ge m$. The solution of (27) does not suggest some input sequence such as u_{k-m-1}, u_{k-m-2}, ..., u_{k-n-1}, and these sequences are needed to perform an experiment. However, these input sequences can be uniquely determined by solving the following:

$$\underset{u_{k-m-1},u_{k-m-2},...,u_{k-n-1}}{\text{Min}} \ \sum_{i=1}^{n-m} (\bar{y}_{k-n+i,} - \tilde{y}_{k-n+i})^2$$

$$\text{s.t.} \quad \bar{y}_{k+1} = \tilde{f}(\bar{y}_k, \bar{y}_{k-1}, ..., \bar{y}_{k-n}, \bar{u}_k ..., \bar{u}_{k-m})$$

$$u_{\min} \le u_{k-i} \le u_{\max}$$

$$|\Delta u_{k-i}| \le \Delta u_{\max} \tag{28}$$

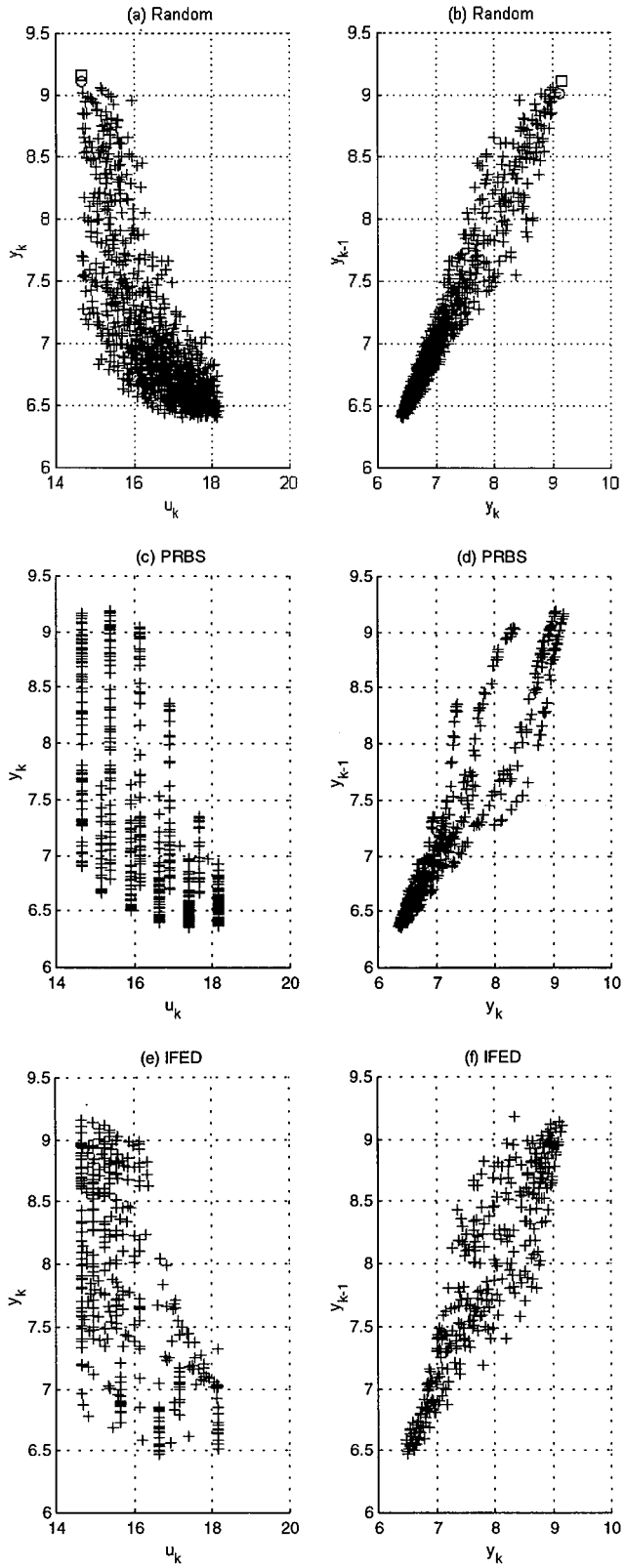

Figure 11. Placements (×) of the experimental data using different approaches: (a, b) random input sequence; (c, d) PRBS signals; (e, f) IFED approach. The symbols (+) are the initial data, and the symbols ○ in a and b are two positions of the plant outputs that appeared in the servo control (set point changed from 7.0 to 9.1) using the random input ANN model.

In this work, (28) is solved using the temporary model $\tilde{f}$ to predict the future. This prediction may not be very accurate until increasing amounts of "efficient" data are added to the temporary training set Ω_0. Figure 3 displays the total computation flowchart.

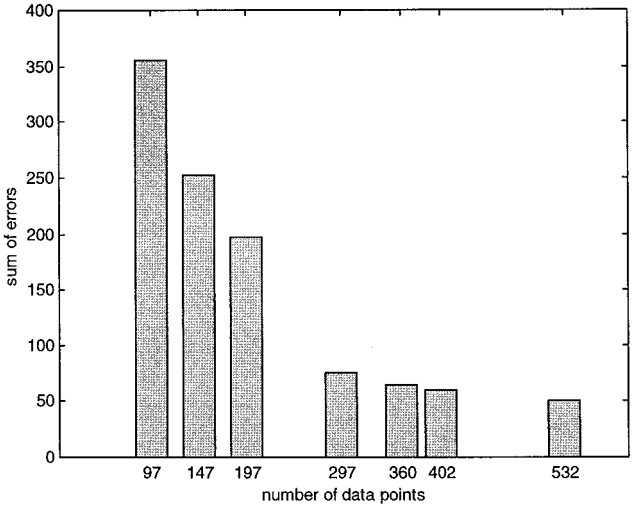

Figure 12. Decrease of the sum of absolute prediction errors of ANN as a function of total number of training data.

5. Illustrative Examples

This section presents three numerical examples. The first example demonstrates the strategy of the information free energy design and compares the differences among the entropy, enthalpy, and free energy designs. In the second example, a system with second order for the output is studied. The operating constraints are then constructed using the Delaunay triangulation approach and compared with the analytical constraints. In the final example, a pH control problem is studied.

5.1. Comparisons among Entropy, Enthalpy, and Free Energy Design. The first example compares the experimental design approaches for an ANN model having the following simple system:

$$y_{k+1} = 3(1 - 3u_k)^2 \exp(-u_k^2 - 16y_k^2) - 10[0.6u_k - 27u_k^3 - 4(y_k - 1)^5] \exp[-9u_k^2 - (4y_k - 1)^2] - \frac{\exp[-(3u_k + 1)^2 - 4(y_k - 1)^2]}{3}$$

Parts a and b of Figure 4 present the 3-D mesh and contour plot of the above dynamic system, respectively. In this case, we approximate the feasible region Φ in (11) by 11 × 11 uniform grid points to y_k and u_k. This results in 121 points in the approximate feasible region of Φ' in (12). The information entropy, enthalpy, and free energy are hence evaluated based on the discretized feasible region Φ'. Assume that a set of initial data exists, as depicted in Figure 5a. According to this figure, pure information entropy design places the experimental points uniformly in the feasible space. The total entropy rapidly increases with an increase of the experimental data, as shown in Figure 5b. As Figure 5c indicates, the total prediction errors of the testing data set using the ANN model decrease. Figure 6a illustrates the experimental design strategy using the information enthalpy (21) only. The experiments are suggested at the places in which the gain changes rapidly, called very nonlinear places, by the information enthalpy. The total enthalpy decreases as the rate in which the experiments are performed increases, as

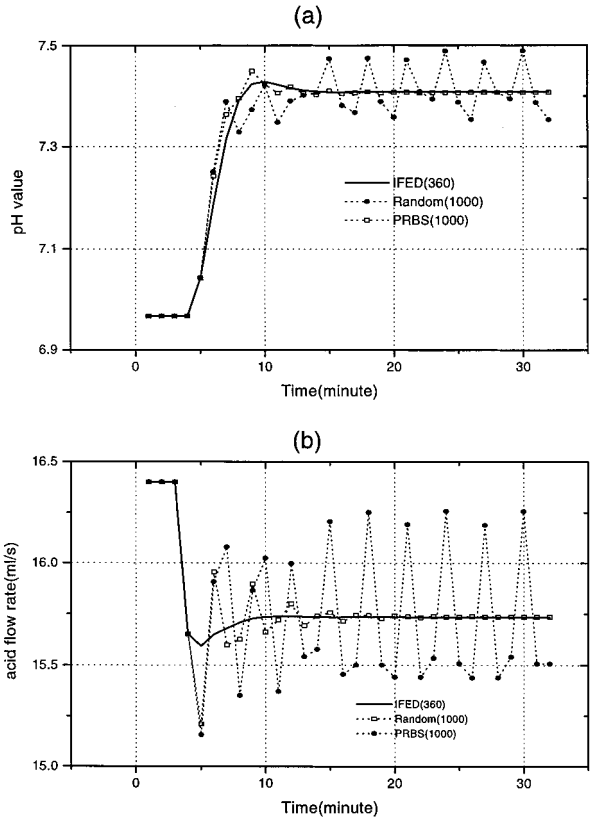

Figure 13. Comparison of the servo behaviors of nonlinear model predictive control based on three different ANN models: (a) pH values vs time; (b) control action vs time. The set point is changed form 7.0 to 7.4.

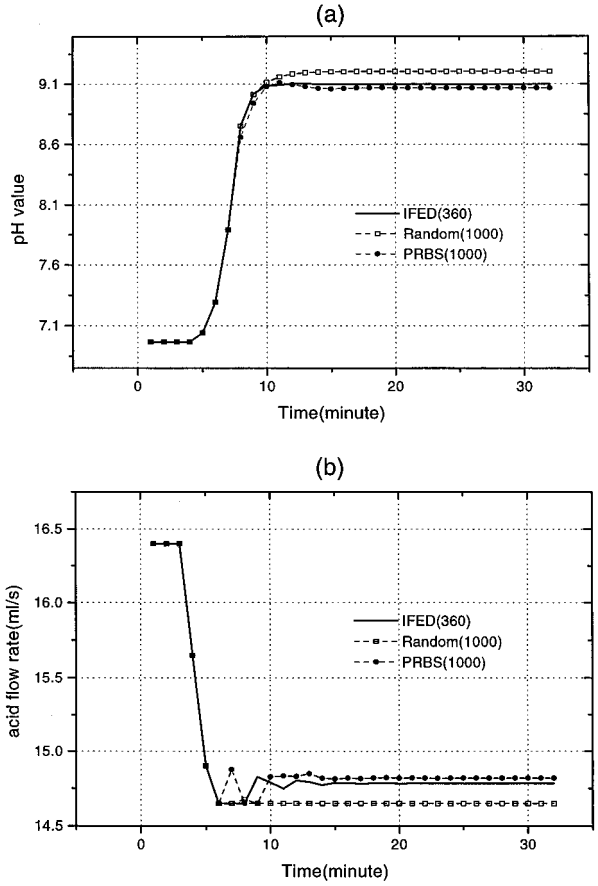

Figure 14. Comparison of the servo behaviors of nonlinear model predictive control based on three different ANN models: (a) pH values vs time; (b) control action vs time. The set point is changed from 7.0 to 9.1.

indicated in Figure 6b. However, as Figure 6c depicts, the information entropy does not decrease as rapidly as the pure information entropy. Figure 6d depicts the decreasing prediction error of the testing data set by the ANN model as a function of experimental data. The error is comparable with the pure information entropy. Figure 7a illustrates the free energy design with a proper selection of information temperature. Parts b−d of Figure 7 depict the decrease of information entropy, enthalpy, and free energy, respectively, with the increase of experimental data. Furthermore, Figure 7e displays the prediction errors of the ANN model that are much better than the pure information entropy and enthalpy. The above prediction errors are obtained based on comparisons between the analytical outputs and ANN outputs on a testing data test with equally spaced 21 × 21 data points.

5.2. Development of a Feasible Region. This section closely examines the following second-order system:

$$y_{k+1} = 0.28{y_{k-1}}^2 + 0.36y_k + u_k/3$$

$$u_k = [0, 1]$$

$$y_k = [0, 1]$$

This and the previous example differ primarily in the need to find the feasible region, the region bounded by (9) and (10). Experiments can be designed based on that region. Figure 8a plots the existing data set, generated by PRBS. The Delaunay approach is hence implemented to approximate the boundaries of the feasible region. Figure 8a compares the analytical boundaries with the approximate boundaries. Based on the approximate feasible region, the information free energy is implemented to obtain the "most" efficient experimental design placements. Parts b−d of Figure 8 compare the approximate boundaries with the actual boundaries for every 10 additional experiments added to the training set. According to those results, the approximate boundaries of the feasible region are improved after 30 iterations of the IFED compared with the real boundaries. Moreover, the ANN model is highly promising in terms of predicting the plant trajectories, as indicated in Figure 9.

5.3. pH Process Modeling and Predictive Control. As Figure 10 depicts, this work simulates a realistic pH control process, as studied by previous researchers.[18] In the case of a sampling time of 1 min, the plant dynamics can be determined as follows:

$$y_{k+1} = f(y_k, y_{k-1}, y_{k-2}, u_k)$$

For comparison, the following ANN models are constructed on the basis of different training data sets:

(1) ANN Model Based on a Training Set Using Random Input Data. One thousand random inputs (acid flow rate) satisfying the constraints, 14.65 mL/s $\leq u_k \leq$ 18.15 mL/s and $\Delta u \leq$ 0.75 mL/s are generated using a random number generator. As assumed herein, the system remains stable at pH 6.9665 initially.

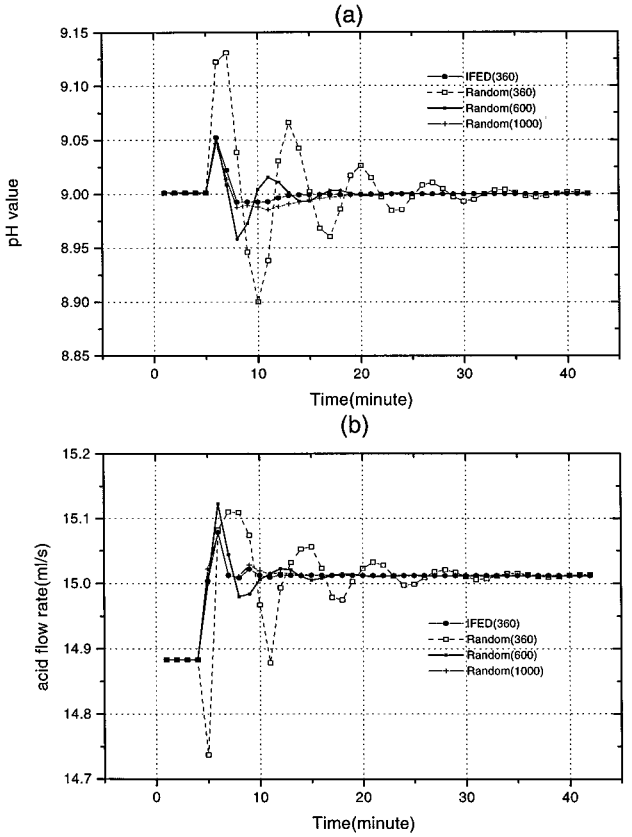

Figure 15. Comparison of the regulation behaviors of nonlinear model predictive control based on four different ANN models: (a) pH values vs time; (b) control action vs time. The flow rate from the buffering tank is changed from 0.55 to 0.3 mL/s.

Hence, the random sequence drives the system. Parts a and b of Figure 11 depict the placements of the experimental data. Three ANN models, with total numbers of 360, 600, and 1000 data points, are trained separately using that data set. We term these three models ANN_{R360}, ANN_{R600}, and ANN_{R1000}, respectively.

(2) ANN Model Based on a Training Set Using PRBS Input Data. Herein, 1000 standard PRBS signals are implemented. The system is also initially assumed to be stable at pH 6.9665. The simulated plant is assumed to be driven by these PRBS signals. Parts c and d of Figure 11 depict the 1000 placements of the PRBS design. An ANN model is trained separately using that data set. We denote the ANN model as ANN_{PRBS}.

(3) ANN Model Based on the IFED Training Set. Assume that a set of initial dynamic data already exists, having a size of PRBS 100 experimental data. After 260 and more iterations of the IFED approaches (totalling 360 and more data points), the sum of errors by the ANN model for the testing data set has been drastically decreased compared with PRBS, as indicated in Figure 12. Figure 12 presents the total absolute errors of a testing set with a total number of 3000 points. According to this figure, the total absolute value rapidly decreases as the number of IFED points accumulates but smoothly decays after the total number of experiments accumulates to 360. Accordingly, we train the dynamic ANN model using only 360 IFED points. We term the ANN model as ANN_{IFED}.

Parts e and f of Figure 11 display the placements of IFED. According to parts a−f of Figure 11, the random input placements fail to drive the system to some

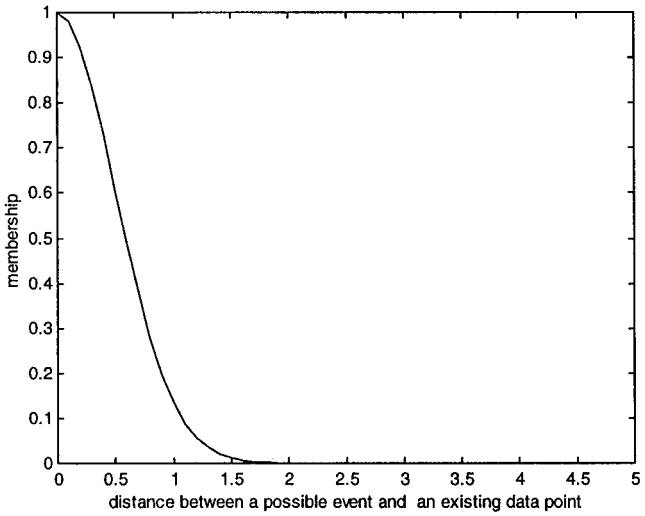

Figure 16. Fuzzy membership of a possible event to an existing data.

portions of the feasible region. The PRBS inputs can drive the system to most portions of the feasible region; however, the experimental points are not distributed as uniformly as IFED. Parts a and b of Figure 13 illustrate the servo behaviors of nonlinear model predictive control, for the case in which the set point of the pH system moves from 6.9665 to 7.4, using the above three ANN models separately. Figure 13a reveals that the ANN_{PRBS} which appeared oscillates around the new set point, while the ANN_{IFED} and random input models behave satisfactorily. Parts a and b of Figure 14 present a case in which the set point of the pH system is moved from 6.9665 to 9.1. In this case, the random input sequence model does not drive the system to the correct set point. This is because the trajectory of the system output was driven to somewhere outside of the training data of the ANN_{R1000} model as shown in Figure 11a,b. ANN_{PRBS} appears much better than the random input sequence model. However, ANN_{IFED} behaves the best among the three ANN models. Parts a and b of Figure 15 also reveal the regulation behavior using the ANN_{IFED} model. Herein, we assume that the flow rate from the buffering tank suddenly changed from 0.55 to 0.3 mL/s and was measured immediately. In this case, in a condition of buffer flow rate of 0.3 mL/s, different ANN's are trained using IFED and random input experimental data. As shown in Figure 15a, the ANN_{IFED} model performs superiorly to other ANN models trained by different numbers of random experiments. Notably, in the case that flow rate from the buffering tank is changed frequently, it is needed to obtain some more ANN models in different flow rates. Next, the control problem (2) using those different ANN models is performed. The regulation control action can thus be obtained by implementing suitable interpolation techniques. However, this is not attempted by this work since it is not the key concern of this work.

6. Conlcusion

This work presents a systematic approach for developing a training data set for a black-box model. The proposed approach is based on the existing data set and a temporary ANN model. In addition, the training data set is used to derive information entropy that measures the uncertainties of the whole feasible region. Informa-

tion enthalpy is also derived to measure the nonlinearity of the entire feasible region. Moreover, new training data are designed based on the compromise between the information entropy and energy−information free energy. The boundaries of the feasible region are also obtained using the concept of the Delaunay approach that is based on the triangulation of the existing data. This work also presents three numerical examples, including a realistic pH control example. The IFED approach is highly promising for all three examples. The ANN model based on the IFED approach can also perform very well for the nonlinear model predictive control of a pH control process.

Acknowledgment

The authors thank the National Science Council of the Republic of China for financially supporting this research under Contract No. NSC87-2622-E007-004.

Notation

c = tuning factor of temperature scheduling
E_{stop} = stopping criteria for the difference between $\mathbf{x}_{pre}$ and $\mathbf{x}_{exp}$
f = dynamic system model
$\tilde{f}$ = temporary ANN model
G = information free energy
ΔG_{stop} = stopping criteria of the free energy change
h = upper constraint of state variables
h' = lower constraint of state variables
H = information enthalpy
M = number of testing data
n = output order of the dynamic system
N = number of experiments
m = input order of the dynamic system
p = probability distribution function
P = number of approximate feasible events
S = information entropy
T = information temperature
T_0 = initial information temperature
u_k = system model input state
Δu = input change
Δu_{max} = upper bound of the input changes
Δu_{min} = lower bound of the input changes
V = Voronoi polygon
$\mathbf{x}$ = augmented event of the dynamic space
$\mathbf{x}_{exp}$ = trajectory performed in eq 28
$\mathbf{x}_{pre}$ = trajectory designed in eq 27
y_k = system model output state
$\tilde{y}$ = system model output state predicted by ANN model
$\bar{y}$ = system model output state designed by IFED
z = tuning factor of temperature scheduling

Greek Symbols

θ = event of the dynamic space
γ = input penalty weighting in the MPC controller
Ψ = testing data set of the black-box model
Ω = training data set of the black-box model
Φ = feasible event set
Φ' = approximate feasible event set
μ = fuzzy membership measure of an event
σ = standard deviation of the Gaussian distribution
σ_μ = standard deviation of the Gausian membership function

Subscripts

k = current time

Appendix: Fuzzy Membership Measure of a Possible Event

Fuzzy mathematics has been a standard technique to determine the "vague" relations among the variables. This work implements the fuzzy membership function to correlate a possible event to an experiment. The shapes of membership functions can be rather arbitrary as suggested by many books.[26] In this work, we suggest implementing the following membership function:

$$\mu(x|x_i) = \mu(x - x_i) = \mu(d) = e^{-|x-x_i|^2/2\sigma_\mu^2}$$

where σ_μ can be roughly determined by the average of the distances among the initial data set Ω_0 such that membership of a middle point between two existing data is not too big or too small. The shape of the membership is demonstrated in Figure 16.

Literature Cited

Bhat, N.; McAvoy, T. J. Use of neural nets for dynamic modeling and control of chemical process systems. *Comput. Chem. Eng.* **1990**, *14* (4/5), 573−583.

Nahas, E. P.; Henson, M. A.; Seborg, D. E. Nonlinear internal model control Strategy for neural network models. *Comput. Chem. Eng.* **1992**, *16* (12), 1039−1057.

Peng, C. Y.; Jang, S. S. Nonlinear rule based model predictive control of chemical processes. *Ind. Eng. Chem. Res.* **1994**, *33* (9), 2140−2150.

Peng, C. Y.; Jang, S. S. Fractal analysis of time-series rule based models and nonlinear model predictive control. *Ind. Eng. Chem. Res.* **1996**, *35*, 2261−2268.

Culter, C. R.; Ramaker, B. L. Dynamic matrix control, a computer control algorithm. Presented at the 86th Meeting of the AICHE Meeting, Houston, April, 1979.

Richalet, J.; Rault, A.; Testud, L.; Papon, J. Model predictive heuristic control: application to industrial process. *Automatica* **1978**, *14*, 413.

Garica, C. E.; Morari, M. Internal model control. 1. A unifying review and some new results. *Ind. Eng. Chem. Res.* **1982**, *21* (2), 308−323.

Jang, S. S.; Joseph, B.; Mukai, H. On-Line Optimization of Constrained Multivariable Chemical Processes. *AIChE J.* **1987**, *33* (1), 26−36.

Hernandez, E.; Arkun, Y. Study of the control-relevant properties of back-propagation neural network models of nonlinear dynamical systems. *Comput. Chem. Eng.* **1992**, *16* (4), 227−240.

Nikolaou, M.; Hanagandi, V. Control of nonlinear dynamical systems modeled by recurrent neural networks. *AIChE J.* **1993**, *39* (11), 1890−1894.

Saint-Donat; Bhat, N.; McAvoy, T. J. Neural net based model predictive control. *Int. J. Control* **1991**, *54* (6), 1453−1468.

Chen, J.; Wong, D. S. H.; Jang, S. S.; Yang, S. L. Product and process development using artificial neural net model and information theory. *AIChE J.* **1998**, *44*, 876−887.

Cavendish, D. A. F.; Frey, W. H. Automatic triangulation of arbitrary planar domains for the finite element method. *Int. J. Numer. Methods Eng.* **1974**, *8*, 679−686.

Weatherill, N. P.; Hassan, O. Efficient 3-dimensional Delaunay triangulation with automatic point creation and imposed boundary constraints. *Int. J. Numer. Methods Eng.* **1994**, *37* (12), 2005−2039.

Subramanian, G.; Raveendra, V. V. S.; Kamath, M. G. Robust boundary triangulation and Delaunay triangulation of arbitrary planar domains. *Int. J. Numer. Methods Eng.* **1994**, *37* (12), 2005−2039.

Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379−423, 623−656.

Deco, G.; Obsradovic, D. *An information theoretic approach to neural computing*; Springer-Verlag: New York, 1995.

Hensen, M. A.; Seborg, D. E. Adaptive nonlinear control of a pH neutralization process. *IEEE Trans. Control Syst. Technol.* **1992**, *2* (3), 169−182.

Gustafsson, T. K.; Waller, K. V. Nonlinear and adaptive control of pH. *Ind. Eng. Chem. Res.* **1992**, *31*, 2681−2693.

Proll, T.; Karim, M. N. Model-predictive pH control using real-time NARX approach. *AIChE J.* **1994**, *40* (2), 269−282.

Pottmann, M.; Seborg, D. E. A radial basis function control strategy and its application to a pH neutralization process. Paper Presented at the Second European Control Conference, *EEC*'93, Groningen, The Netherlands, 1993.

Rhodes, C.; Morari, M. The false nearest neighbors algorithm: An Overview. *Comput. Chem. Eng.* **1997**, *21*, S1149−1154.

Jang, J.-S. R.; Sun, C. T.; Mizutani, E. *Neuro-Fuzzy Soft Comput.* **1996**, 180−186.

Goodwin, G. C.; Payne, R. L. *Dynamic system identification: experiment design and data analysis*; Academic Press: New York, 1977.

MathWorks. Delaunay function, *MATLAB user manual*; MathWorks: Natick, MA, 1997.

Zimmermann, H. J. *Fuzzy Set Theory and Its applictaions*; Kluwer-Nijhoff: Dordrecht, The Netherlands, 1985.

Received for review February 24, 1998
Accepted June 3, 1998

IE9801249