

PROCESS DESIGN AND CONTROL

Generalized Multivariable Dynamic Artificial Neural Network Modeling for Chemical Processes

Jun-Shien Lin, Shi-Shang Jang,* Shyan-Shu Shieh,[†] and M. Subramaniam

Chemical Engineering Department, National Tsing-Hua University, Hsin-Chu 300, Taiwan

This work presents a novel systematic approach to acquire good-quality plant data that can be efficiently used to build a complete dynamic empirical model along with the use of partial plant knowledge. A generalized Delaunay triangulation scheme is then implemented to find feasible operating boundaries that may be nonconvex on the basis of the existing plant data. The Akaike information index is adopted to assess partial plant knowledge as well as noisy plant data. The information free energy is calculated for acquisition of good-quality new plant data that will improve the dynamic model. The new experimental data suggested by the information analysis, together with the previous data and prior plant knowledge, are used to train a new dynamic empirical model. Multivariable model predictive control for a high-purity distillation column using the acquired model based on the proposed approach is also studied. Comparing with PRBS and RAS schemes, the proposed approach outperforms the rest.

1. Introduction

Model predictive control (MPC) is becoming a standard recently in the chemical industries. Tremendous progress has been made theoretically¹ and practically.² However, only linear MPC is available in the process industry, although many researchers have investigated the possibility of implementing a nonlinear empirical model, such as an artificial neural network (ANN) model.^{3–5} This study tries to address the following difficulties caused due to implementing nonlinear MPC.

In the case of a multivariable system, too many data are required to train a “complete” ANN model. Unfortunately, no systematic approach exists to develop a “complete” nonlinear model.

It is undesirable for an empirical model to perform extrapolation, and also there is no guarantee that an event of the on-line model prediction always falls into the region described by the training set.

On-line measurements are contaminated by signal noise. When a training set is not large enough, the prediction capability of an empirical model would be impaired by noise-contaminated data.

As partial plant knowledge is available, it should be possible to include partial plant knowledge with plant data together so that the number of needed experiments can substantially be reduced.

In the past decade, many researchers have studied to incorporate plant knowledge into a black box model. Thompson and Kramer⁶ combine a physical model in parallel with a radial based neural network to enhance the extrapolation ability of the empirical model. Psi-

chogios and Ungar⁷ implement an ANN in series with a physical model to estimate the parameters of the first-principle model. The author⁸ applies a physical model to extend the domain of a training set. The incorporation of partial plant knowledge into a black box model is termed as a “gray box” model. Van Can^{9,10} has applied gray box models in chemical process systems. However, all these works utilize a first-principle model that is usually too expensive and nearly impossible to obtain in industrial scale processes.

Akaike information was originally proposed as a regression index.^{11,12} The minimization of Akaike information criteria (AIC) has been widely implemented on model diagnosis and model structure selection.^{13,14} The physical meaning of AIC is a compromise between regression error and the degree of freedom of a regression model. Conditional Akaike information criteria (CAIC) are used in this work to determine the optimal combination of operating data and several types of prior knowledge. A neural network model that trains with these data together is derived. We also consider that the data obtained from different sources should be weighted differently. These weights can be determined by optimizing the CAIC.

Black box models, for example, ANN models, are usually derived using plant data. Since ANN models are parameter-rich, they may be overfitted if a training data set is not large enough. When applied to chemical processes, the prediction capability of ANN models would be further impaired by the noise-contaminated plant data. In other words, overfitting of the models becomes much more severe when training data are noisy. Dong and McAvoy¹⁵ developed a principal neural network, and Johansen¹⁶ also derived a spline regression to treat noisy data. In this work, we introduced the

* Corresponding author. Telephone: +886-3-571-3697. Fax: +886-3-571-5408. E-mail: ssjang@che.nthu.edu.tw.

[†] Current address: Department of Occupational Safety and Hygiene, Chang Jung University, Tainan County, Taiwan.

idea of smoothing plant data to ANN model training to ease the noise problem.

On-line applications of black box models must ensure that the models do not perform extrapolation. In this work, we term the "completeness" of a model as an empirical model that always performs interpolation during predicting any new event. Note that interpolation/extrapolation are defined as points lying inside/outside the nonconvex hull throughout the paper. If a point lies inside the nonconvex hull of a data set, the model prediction at this point is a result of interpolation. Until now, very limited works have focused on this issue. Raju and Cooney¹⁷ proposed to improve the model by conducting new experiments that are determined by reviewing training set data. Our previous works^{18,19} proposed that the model should determine new experiments to ensure the accuracy of its prediction by investigating the total information free energy of the training set. Assuming the convexity of the operation boundaries, Lin and Jang (1998) proposed to implement Delaunay triangulation²⁰ to find the boundaries of feasible operation events. This is not very true in highly nonlinear systems such as high-purity distillation columns. In this work, one step further, we developed the generalized Delaunay triangulation approach to locate the nonconvex operation region of a nonlinear system. The information analysis based on the nonconvex operation region guarantees the "completeness" of an empirical model. This implies that an empirical model always performs interpolations only.

Information entropy is derived by Shannon²¹ to evaluate the uncertainty of a random variable. Recently, substantial applications of this theory can be found in the literature.²² In contrast to information entropy, the authors derived information enthalpy to evaluate the nonlinearity of a system.^{18,19} We suggest that it is more appropriate to create new experiments by minimizing the information free energy that is a function of the information entropy and the information enthalpy.

The Taguchi method²³ is a widely used technique for experimental design in the industry. Our proposed scheme is different from the Taguchi method in the following aspects. The new experimental points in our work are based on the minimization of the information free energy by varying the location of new data points while, in the Taguchi method, they are based on an orthogonal table and maximization of the signal-to-noise ratio. The Taguchi method only suggests an individual manipulated variable operating at a predetermined level. The proposed technique takes both manipulated variables and response variables into consideration before determining a new experimental point. Moreover, the proposed algorithm is an evolutionary improvement mechanism while the Taguchi method is not. Therefore, the proposed technique can be applied to a dynamic system while the Taguchi method can only be used in a steady-state system.

Model predictive control of high-purity distillation columns has been a challenging problem in the area of process control due to its nonlinearity and multivariable nature. Neural network models are an active approach for the MPC of distillation systems. Baratti et al.²⁴ developed a dynamic empirical neural network model for top composition control of a multicomponent distillation column. Savkovicstevanovic³ developed inverse models for single-variable control of a distillation system. Multivariable systems are seldom discussed due

to the complexity of the data structure. Doyle III and Shaw²⁵ discussed a multivariable recurrent dynamic neuron network in MPC for a high-purity distillation column. Their approach, however, did not guarantee that the empirical model does not perform extrapolation during on-line operation. In this work, an empirical multivariable model that incorporates partial plant knowledge of a high-purity distillation column is developed. The neural network model is "complete" and hence will not perform extrapolation in plant operation. The MPC based on the empirical model works satisfactorily for the dual-temperature control of the high-purity distillation column.

The objective of this work is to develop a systematic approach for a dynamic multivariable ANN model that is robust and complete to use in MPC. The contributions of this work can be summarized as follows:

A new training policy for ANN models is developed. This policy is very useful in the case that data are noisy and/or partial plant knowledge is available.

An efficient experimental design algorithm is derived for building up an empirical dynamic model. The approach identifies the feasible operation region for a dynamic system so that the empirical model does not perform extrapolation.

A novel approach to obtain the nonconvex feasible region for a set of data is proposed.

In this paper, theoretical development of a data-oriented model is discussed in the next section. The details of the algorithm are presented in section 3. Three illustrated examples including a realistic multicomponent distillation column are given in section 4. Conclusive remarks are given in section 5.

2. Theory

Consider the following generalized lumped system:

$$\begin{aligned}\dot{\mathbf{x}} &= \mathbf{f}(\mathbf{x}, \mathbf{m}) \\ \mathbf{y} &= \mathbf{g}(\mathbf{x}, \mathbf{m})\end{aligned}\quad (1)$$

where $\mathbf{x} \in R^n$ are state variables, $\mathbf{m} \in R^m$ are system inputs, and $\mathbf{y} \in R^p$ are system outputs or the on-line measurements of a plant. The functions \mathbf{f} and \mathbf{g} are presumably unknown due to lack of plant knowledge. Assume that the above system can be described by the following discrete time system with a determined sampling time T :

$$\mathbf{y}_{k+1} = \mathbf{h}(\mathbf{y}_k, \mathbf{y}_{k-1}, \dots, \mathbf{y}_{k-s}, \mathbf{m}_k, \dots, \mathbf{m}_{k-s'}) \quad (2)$$

where k is the current time. For simplicity of formulation, let us take $m = p = 2$, that is, a 2×2 system.

$$\mathbf{y} = \begin{bmatrix} y^1 \\ y^2 \end{bmatrix}, \quad \mathbf{m} = \begin{bmatrix} m^1 \\ m^2 \end{bmatrix}$$

The following derivation is general to all multiinput–multioutput (MIMO) systems.

$$\begin{aligned}y_{k+1}^1 &= h_1(y_k^1, \dots, y_{k-n1}^1, y_k^2, \dots, y_{k-n2}^2, m_k^1, \dots, m_{k-m1}^1, m_k^2, \dots, m_{k-m2}^2) \\ y_{k+1}^2 &= h_2(y_k^1, \dots, y_{k-n3}^1, y_k^2, \dots, y_{k-n4}^2, m_k^1, \dots, m_{k-m3}^1, m_k^2, \dots, m_{k-m4}^2)\end{aligned}\quad (3)$$

It should be noted that the determinations of the sampling time T and the system orders $n1, n2, n3, n4, m1, m2, m3,$ and $m4$ are nontrivial problems. Many studies can be found in the literature.²⁶⁻²⁸ Herein, we assume that system orders and sampling time for a specific system can be obtained.

Denote an event that happened to system 1 or 2

$$\phi = (y_{k'}^1, \dots, y_{k-n5}^1, y_{k'}^2, \dots, y_{k-n6}^2, m_{k'}^1, \dots, m_{k-m5}^1, m_{k'}^2, \dots, m_{k-m6}^2)$$

where $m5 = \sup(m1, m3), m6 = \sup(m2, m4), n5 = \sup(n1, n3), n6 = \sup(n2, n4),$ and

$$\mathbf{y}_{k+1} = \begin{bmatrix} y_{k+1}^1 \\ y_{k+1}^2 \end{bmatrix} = \mathbf{h}(\phi)$$

and the feasible region of system 1 or 2 is then represented by

$$\Phi = \{\phi^i | \phi^i = \text{all possible events}, i = 1, 2, \dots\}$$

2.1. Problem Formulations. Consider a system 1 or 2; there exists an experimental event set

$$\Omega = \{\omega^i | \omega^i, i = 1, \dots, N\} \subset \Phi$$

such that for all events

$$\omega = [(\bar{y}_{k'}^1, \dots, \bar{y}_{k-n5}^1, \bar{y}_{k'}^2, \dots, \bar{y}_{k-n6}^2, m_{k'}^1, \dots, m_{k-m5}^1, m_{k'}^2, \dots, m_{k-m6}^2)] \in \Omega$$

$\bar{\mathbf{y}}_{k+1} = \mathbf{h}(\Omega)$ are determined by experiments and N is the size of the of experimental event set. $\bar{\mathbf{y}}_k$ is on-line measurement and is contaminated by white noise. Let new experiments be available; therefore, after new experiments are performed, the size of Ω is expanded.

Partial knowledge of the system is assumed to be the event set $\Psi \subset \Phi,$ such that for

$$\psi^i = [(\widehat{y}_{k'}^1, \dots, \widehat{y}_{k-n5}^1, \widehat{y}_{k'}^2, \dots, \widehat{y}_{k-n6}^2, m_{k'}^1, \dots, m_{k-m5}^1, m_{k'}^2, \dots, m_{k-m6}^2)] \in \Psi$$

$$\widehat{\mathbf{y}}_{k+1} = \begin{bmatrix} \widehat{\mathbf{y}}_{k+1}^1 \\ \widehat{\mathbf{y}}_{k+1}^2 \end{bmatrix} = \mathbf{h}'(\Psi)$$

\mathbf{h}' is known. Here we assume that the output of Ψ is not noisy but that the knowledge can be incorrect, that is, $\mathbf{h}' \neq \mathbf{h}.$ The purpose of this work is to find the following model:

$$\begin{bmatrix} \hat{y}_{k+1}^1 \\ \hat{y}_{k+1}^2 \end{bmatrix} = [\text{ANN}(y_{k'}^1, \dots, y_{k-n5}^1, y_{k'}^2, \dots, y_{k-n6}^2, m_{k'}^1, \dots, m_{k-m5}^1, m_{k'}^2, \dots, m_{k-m6}^2)] \cong \mathbf{h}(\Phi) \quad (4)$$

based on the training set $\Omega \cup \Psi,$ where the elements of Ω are from on-line experiments while the elements of Ψ are from partial knowledge.

The new approach developed in this work is to treat the above two different data sets by the following philosophy:

(i) There must exist suitable penalties for the use of operating data and partial knowledge in model training.

The penalties can be adjusted on the basis of their significance.

(ii) The on-line data are usually noisy, while, in most cases, the partial plant knowledge is not noisy. It would be better to smooth the operating data before both sets can be put together to train the ANN models.

The details of the above approach are presented in the next section, while the development of the extended experimental set is discussed in section 3.1.

2.2. Development of the Dynamic ANN Models Using Operating Data and Partial Knowledge.

Given an ANN model in eq 4 as a three-layer (P input nodes, H hidden nodes, and 2 output nodes) feed-forward ANN model, denote the following input layer for the model,

$$\chi = (\chi_1, \chi_2, \dots, \chi_P) = (y_{k'}^1, \dots, y_{k-n5}^1, y_{k'}^2, \dots, y_{k-n6}^2, m_{k'}^1, \dots, m_{k-m5}^1, m_{k'}^2, \dots, m_{k-m6}^2)$$

where $P = n5 + n6 + m5 + m6 + 4;$ then, the output of the i th neuron of the hidden layer is

$$h_i(\chi) = \tanh\left(\sum_{j=1}^P w_{ij}^h \chi_j + b_i^h\right)$$

$$w_{ij}^h \in R^H \times R_P, b_i^h \in R^H \quad (5)$$

where $i = 1, \dots, H.$ The l th output of the ANN model is

$$\hat{y}_{k+1}^l = \sum_{i=1}^H w_{li}^o h_i(\chi) + b^o$$

$$w_{li}^o \in R^2 \times R^H, b^o \in R^2 \quad (6)$$

where $l = [1, 2].$ w_{ij}^h and b_i^h in eq 5 and w_{li}^o and b^o in eq 6 are determined by minimizing the following objective function:

$$J(\mathbf{w}) = \sum_{i=1}^N |\hat{\mathbf{y}}_{k+1}(\omega_i, \mathbf{w}) - \bar{\mathbf{y}}_{k+1}(\omega_i)| + \lambda \sum_{i=1}^{N_G} \left\| \frac{\partial^2 \hat{\mathbf{y}}(\omega_i)}{\partial \omega_i^2} \right\|^2 + \beta \sum_{i=1}^R |\hat{\mathbf{y}}_{k+1} - \hat{\mathbf{y}}_{k+1}(\psi_i)| \quad (7)$$

where \mathbf{w} is $\{w_{ij}^h, b_i^h, w_{ij}^o, b^o | j = 1, \dots, H; i = 1, \dots, P; l = 1, 2\},$ N_G is the number of grid points where the smoothness factor is calculated, and R is the number of partial knowledge data points.

The first term in eq 7 is fitting the operating data, the second term is smoothing the noisy operating data, and the third term is fitting the partial knowledge for ANN's prediction. The penalty factors λ and β are determined by the following derivation.

Given N total plant data, which can be separated into two sets (that is, one is the training set (K data), and the other is the validation set ($N-K$)), we determine the minimum of the following conditional AIC¹⁴

$$\min_{\zeta} V(\zeta) = \ln \left[\frac{1}{N-K} \sum_{i=1}^{N-K} \left| \hat{\mathbf{y}}_{k+1}(\omega_i, \zeta) - \bar{\mathbf{y}}(\omega_i, \zeta) \right|^2 \right] + \frac{2}{N-K} \dim[\zeta, \mathbf{w}] \quad (8)$$

by manipulating $\zeta = (\lambda, \beta).$

It should be noted that the smoothness term

$$\sum_{i=1}^{N_G} \left\| \frac{\partial^2 \hat{\mathbf{y}}(\omega_i)}{\partial \omega_i} \right\|^2$$

in eq 7 can be analytically derived as the following:

$$\frac{\partial^2 \hat{\mathbf{y}}_{k+1}(x^p)}{\partial (x^p)^2} = \sum_{i=1}^2 \mathbf{w}^{h^T} \frac{-2}{\text{diag}[\csc h(\mathbf{w}^h x^p + \mathbf{b}^h)]^2 \text{diag}[\tanh(\mathbf{w}^h x^p + \mathbf{b}^h)] \cdot \text{diag}[\mathbf{w}_i^q] \cdot \mathbf{w}^h} \times \quad (9)$$

2.3. Approximation of the Nonconvex Feasible Region. In most cases, the accuracy of the testing set does not automatically satisfy the on-line performance of the ANN models. Even worse, extrapolation events may happen during the on-line operation. It is hence very desirable to find the feasible event set Φ , or some other set Φ' which is very close to Φ . Let us first denote a convex hull $C(\Omega)^{29}$ for a set of dynamic events Ω . Traditional interpolation (i.e. for a convex hull) could be ensured if $C(\Omega) = \Phi' \approx \Phi$.

Definition 2-1 (Convex Hull). Given $\Omega = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, then the convex hull of Ω is the smallest convex set containing Ω , denoted as $C(\Omega)$. $C(\Omega)$ is bounded by a set of linear inequalities that could be determined uniquely as follows.

$$\min_{A,b} \text{area}(C(\Omega)) \text{ s.t. } C(\Omega) = \{x_j \in \Omega | Ax_j \leq b\} \quad (10)$$

Definition 2-2 (Delaunay Triangulation). Given $\Omega = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, Delaunay triangulation is a set of lines connecting an individual point to its most close neighbors of the region. If edges used by more than two triangles are canceled, then the boundary of the convex hull appears.

However, it should be noted that the on-line data are contaminated by white noise, and the convergence of a convex hull is shown by the following proposition:

Proposition 2-1 (Convergence of Convex Hull of Noisy Data). (a) Given $D = \{d_1, \dots, d_N\}$ as a Gaussian normal process (white noise), $C(D)$ converges. (b) Given $\Omega = \{\omega_1, \dots, \omega_N\}$ as the set of dynamic events of a deterministic process and $\Omega' = \{\chi_i = d_i + \omega_i | i = 1, \dots, N, d_i \in D, \omega_i \in \Omega\}$, then $C(\Omega')$ converges.

Proof. See Appendix.

The major prediction error in modeling is usually caused by extrapolation. Since $C(\Omega')$ is available, this study proposes to evaluate $C(\Omega')$ over the full data structure instead of a testing set. In case Φ' is convex, the approach is presented in our previous paper.¹⁹ However, many cases of chemical systems, such as high purity columns, are nonconvex. We present the following theoretical basis for the nonconvex system.

Definition 2-3 (Nonconvex Hull). Given $\Omega' = \{\chi_1, \dots, \chi_N\}$, Divide Ω' into an m subset $\Omega' = \Omega'_1 \cup \Omega'_2 \cup \dots \cup \Omega'_m$. The nonconvex hull, $NC(\Omega') = \cup_{i=1}^m C(\Omega'_i)$ can be decided as follows.

$$\min \bigcup_{i=1}^m \text{area}[C(\Omega'_i)] \text{ s.t. } \bigcup_{i=1}^m C(\Omega'_i) \text{ is continuous} \quad (11)$$

Definition 2-4 (Generalized Delaunay Triangulation). Generalized Delaunay triangulation is a set of

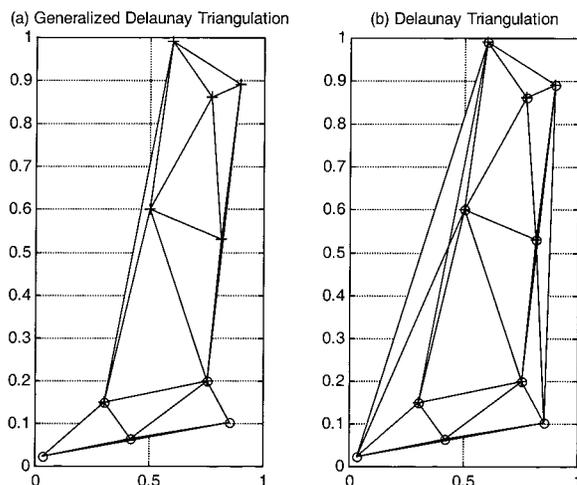


Figure 1. Generalized Delaunay triangulation and Delaunay triangulation.

lines connecting an individual point to its most close neighbors for points in Ω'_j . If edges used by more than two triangles and line segments overlapped with $\cup_{i=1}^m C(\Omega'_{i \neq j})$ are canceled, then the boundary of the nonconvex hull appears.

Nonconvex hull computing is processed as the following steps

Step 1. Set $m = 1$.

Step 2. Divide $\Omega' = \{\omega_1, \omega_2, \dots, \omega_N\}$ into m subsets $\Omega'_1, \Omega'_2, \dots, \Omega'_m$; then the possible combinations for N events belonging to m sets is (N, m) .

Step 3. Perform Delaunay triangulation with respect to $\chi_i, \chi_i \in \Omega'_i$.

Step 4. Cancel redundant line segments and edges defined in definition 2-4 and acquire the nonconvex hull.

Step 5. Calculate the area of the nonconvex hull.

Step 6. Repeat steps 2–5 until all possible combinations of subset formation are tried.

Step 7. Find proper nonconvex hull with minimal area.

Step 8. Set $m = m + 1$.

Step 9. Stop until the nonconvex area no longer decreases; otherwise, go to step 2.

Figure 1 demonstrates the difference between Delaunay and generalized Delaunay triangulation approaches applied to a set of data. Initial data are divided into two subsets denoted as + and \circ in Figure 1a. The original Delaunay triangulation is shown in Figure 1b. Generalized Delaunay triangulation gives a smaller compact hull, as shown in Figure 1a, than the original Delaunay triangulation approach shown in Figure 1b in the case that the smallest convex hull is nonconvex.

The following proposition is a natural extension of proposition 2-1:

Proposition 2-2 (Convergence of Nonconvex Hull of Noisy Data). Given $D = \{d_1, \dots, d_N\}$ as white noise, and $\Omega = \{\omega_1, \dots, \omega_N\}$ is the dynamic event set of a deterministic process. The operating data are defined as follows:

$$\Omega' = \{\chi_i = d_i + \omega_i | i = 1, \dots, N; d_i \in D; \omega_i \in \Omega\}$$

The nonconvex hull $NC(\Omega')$ converges.

Proof. See Appendix.

The above proposition remains true when partial knowledge is present, that is $\Phi \approx \Phi' = C(\Omega' \cup \Psi)$.

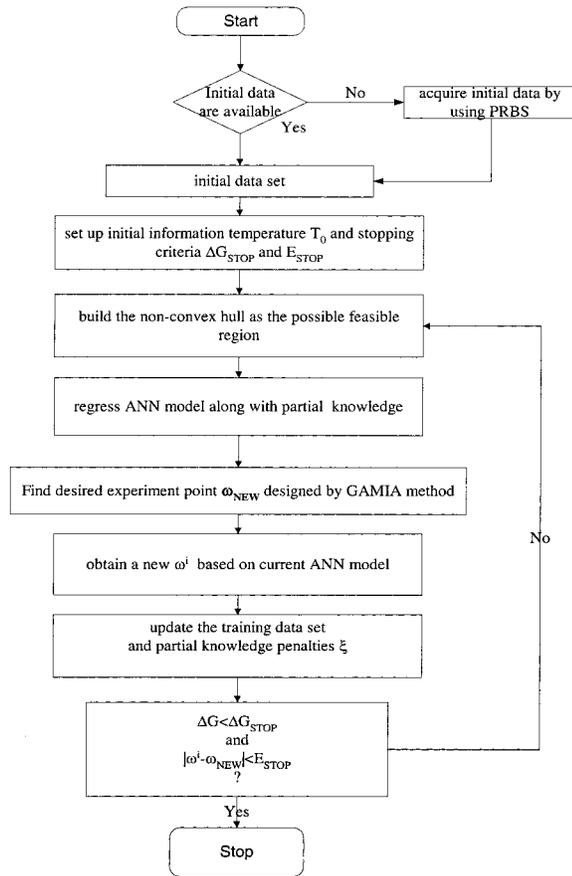


Figure 2. Flowchart of GAMIA.

3. Algorithm

3.1. Development of Extended Experimental Data Set. Given the original training set Ω_0 and assuming that Φ' can be approximated by the above nonconvex hull $NC(\Omega_0)$, our purpose is to develop an extended training set $\Omega_1, \Omega_2, \dots, \Omega_K$, such that

(i) the nonconvex hull N is bounded by the total training set

$$\Omega' = \Omega_0 \cup \Omega_1 \cup \dots \cup \Omega_K \cup \Psi, NC(\Omega') \approx \Phi'$$

(ii) for all $\omega \in \Phi'$,

$$\hat{\mathbf{y}}_{k+1} \approx \mathbf{y}_{k+1}$$

The first condition guarantees that the model does not perform extrapolation while the second condition is to ensure the prediction quality of the ANN models. To meet the above requirements, the authors developed an information free energy experimental design (IFED) in our previous work¹⁹ for convex hull systems. This work extends the concept of IFED from the convex feasible region and SISO systems to nonconvex and MIMO systems. Note that Φ' is denoted as the nonconvex hull and is a continuous set with basically infinite elements. However, the following information properties are denoted to a discretized set with a finite size. Define $\Phi'' = \{\omega_1, \omega_2, \dots, \omega_Q\}$ to be a discretized approximation of Φ' with all vertexes of Φ' .

(1) Information Entropy. The first objective of the placement of a new experimental point is to put the new data at a region where the original data are lacking. Assume the probability measure of an event $\omega \in \Phi''$ can

be determined by all events $\omega_i \in \Omega \cup \Psi, i = 1, \dots, N$, as the following:

$$p(u) = \frac{1}{\sqrt{2\pi\sigma}} e^{-|u|^2/2\sigma^2} \quad (12)$$

where $u \in [-\infty, \infty]$, and $\sigma^2 = 1/\sum_{i=1}^N \mu(\omega|\omega_i)$, with $\mu(\omega|\omega_i) = \mu(d = |\omega - \omega_i|)$ denoting the fuzzy membership of a possible event $\omega \in \Phi'$ belonging to a training data $\omega_i \in \Omega \cup \Psi$. Later, evaluate the following information entropy:³⁰

$$S(\omega) = \int_{-\infty}^{\infty} p(u) \ln p(u) du = -\frac{\ln(2\pi\sigma^2)}{2} \quad (13)$$

The total information entropy of Φ'' is hence denoted as

$$S = \sum_{j=1}^Q S(\omega_j) \quad (14)$$

According to eq 14, the change of the total information entropy of Φ'' for a new experiment ω added to the training set can be evaluated. If the experiment is added to the place where the training set is more sparse, then the total entropy is lowered.

(2) Information Enthalpy. The second objective of the placement of a new experimental data is to get more information where the system nonlinearity is severe. Assume that j iterations of experimental design have been performed; that is, $\Omega_1, \Omega_2, \dots, \Omega_j$ new experimental sets are developed. Given MIMO system 2 and \hat{y}_{k+1}^j as the i -th output of ANN models for eq 2 based on the training set $\Omega_0 \cup \Omega_1 \cup \dots \cup \Omega_j \cup \Psi$, then for any event $\omega \in \Phi''$

$$H(\omega) =$$

$$\sum_{i=1}^2 \text{norm} \left| \begin{array}{cccc} \frac{\partial^2 \hat{y}_{k+1}^j}{(\partial y_k^1)^2} & \frac{\partial^2 \hat{y}_{k+1}^j}{\partial y_k^1 \partial y_{k-1}^1} & \dots & \dots \\ \frac{\partial^2 \hat{y}_{k+1}^j}{\partial y_k^1 \partial y_{k-1}^1} & \frac{\partial^2 \hat{y}_{k+1}^j}{(\partial y_{k-1}^1)^2} & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \frac{\partial^2 \hat{y}_{k+1}^j}{(\partial m_{k-m6}^1)^2} \end{array} \right| \left\| \left(1 + \frac{1}{\sigma^2} \right) \right. \quad (15)$$

where σ^2 is denoted the same with eq 12. The total information enthalpy of the training set Φ'' can be evaluated:

$$H = \sum_{i=1}^Q H(\omega_i) \quad (16)$$

According to the above definition, the information enthalpy eq 16 can be evaluated. A new experimental datum added to the training set will lower the total information enthalpy if it is placed at a more nonlinear place.

(3) Information Free Energy. In this work, the above needs—information entropy and information enthalpy—are balanced by introducing a new informa-

tion property—information free energy. At the beginning, we assume that the ANN models are inaccurate due to the lack of experiments. It is more important to put the new experiments at the places where data are sparse. As more experiments accumulate, the model becomes more accurate; then it is more important to put the data at a region where the system is more nonlinear. To perform this compromise, we introduce the following scaling factor “information temperature” and “information free energy”:

$$G = H - TS \quad (17)$$

where temperature T is changed as a function of number of experiments:

$$T(N) = T_0 \exp(-cN^{1/z}) \quad (18)$$

where N denotes the total number of experiments including the initial data set, and T_0 and c are tuning parameters.¹⁹

3.2. Flow Chart. Figure 2 shows the algorithm developed in the previous section. Assume that the noisy operating plant data set Ω_0 and partial knowledge already exist. First of all, the parameters of the generalized objective function eq 7 of training of ANN models are determined by solving eq 8. Next, the nonconvex hull is implemented to approximate the feasible region. The approximate feasible event set Φ' is, in turn, discretized and implemented to evaluate the information properties. The extended experimental data set is developed to minimize the information free energy. These experiments are performed and go back to check for the nonconvex hull. However, if the total number of experiments is accumulated to a predetermined level, the problem of conditional AIC is solved again to determine the penalties of partial knowledge and the smooth factor. The algorithm presented in Figure 2 is termed Generalized Artificial Neural Network Modeling Based on Information Analysis (GAMIA for short). This terminology will be used in the next section for convenience.

4. Examples

In this section, three examples to demonstrate this novel approach are presented. In the first example, we only emphasize the importance of the smoothness factor in training an ANN model, as the size of the experimental data set is small and it is noisy. The second example is the modeling of a simple nonlinear dynamic process. In this example, the importance of conditional AIC is fully demonstrated. In the third example, a simulated multicomponent distillation column is presented. Peng and Jang²⁶ found that satisfactory performance on the system could not be achieved by conventional feedback control. The feasible region of this dynamic system is nonconvex. The ANN model for this dynamic system is obtained by the GAMIA algorithm and implemented for MPC. The results show that our approach is valid and useful for multivariable model predictive control on this complex system.

4.1. Example 1: ANN Modeling of a Simple System. Consider the following quadratic system:

$$y = x_1^2 + x_2^2 + \epsilon \quad (19)$$

where $\epsilon \in N(0,2)$. Our objective is to build an ANN model

using 16 experimental data points which are generated by setting $x_1 = -3, -1, 1, 3$ and $x_2 = -3, -1, 1, 3$. The maximum number of nodes of the hidden layer is set to be 4 when considering the number of fitting parameters in the ANN model. For the sake of comparison, two types of training strategies are used for all cases. The first training strategy uses the first term of eq 7, and this is the conventional ANN training strategy. The second strategy takes the smoothness factor and/or partial knowledge into account, that is, the second and third terms of eq 7. The parameter λ is determined by solving the conditional AIC criteria of eq 8. However, this example excludes the partial knowledge term. The comparisons of the two models are displayed in Figures 3. The figures show the contour plots with and without the smoothness term in training ANN models with a 4 nodes hidden layer. Equation 19 is a simple circle; the contour plot with the smoothness factor is more accurate than the case without the smoothness factor. To demonstrate the prediction capability of all these models in this example, the sum of the absolute errors is calculated. With a testing set of 961 points, the sum of absolute errors in the case without the smoothness factor is 1259.8, which is twice the magnitude of that for the case with the smoothness factor, that is, 639.4. In this example, the results conclude that the smoothness factor dramatically improves model accuracy, especially in the environment of noisy measurements.

4.2. Example 2: Second-Order Dynamic System. Assume an unknown system that is defined by the following dynamic equation:

$$y_{k+1} = 0.28y_{k-1}^2 + 0.36y_k + u_k/3 \quad (20)$$

where $u_k \in [0,1]$ and $y_k \in [0,1]$. We proposed its steady-state equation is assumed and it is noise free as

$$0.28y_s^2 - 0.64y_s + u_s/3 = 0 \quad (21)$$

The above equation represents the prior knowledge that is partially correct. It is our purpose to obtain the following dynamic model based on a set of noisy operating data Ω_0 .

$$y_{k+1} = \text{ANN}(y_k, y_{k-1}, u_k) \quad (22)$$

Using the PRBS scheme,³¹ we sample 20 points as the initial data set Ω_0 . The data are contaminated by a normal distributed noise $N(0,0.02)$. A set of 41 points obtained from eq 21 represents partial knowledge data, Ψ . To enhance the training set, the GAMIA algorithm is implemented to acquire additional data. All three different data sets, namely initial data, partial knowledge data, and GAMIA data, denoted by (+), (O), and (x), respectively, are plotted in Figure 4. A 3D plot is shown in Figure 4a while its projection into the y_k and y_{k-1} plane is shown in Figure 4b. The solid line in the figure is obtained by solving eq 20 analytically. The approximate boundary of the feasible region obtained by conducting the convex hull scheme is shown as dotted lines. Observing the figure closely, we find that all GAMIA data are almost located at the void place where the initial data and the partial knowledge data rarely appear.

To investigate the effect of the smoothness factor in ANN modeling, we take three lots of the initial data set, that is, 30, 60, and 100. Since the total feasible region

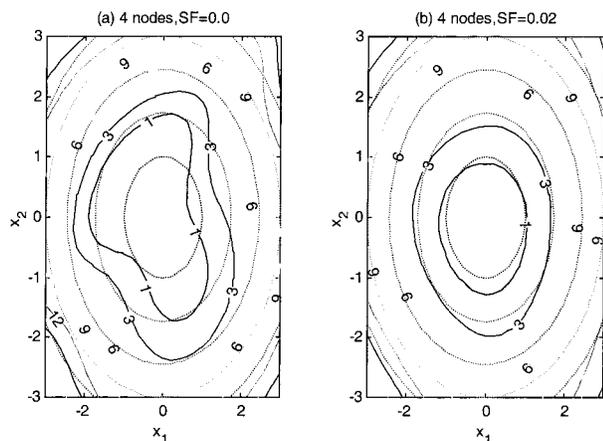


Figure 3. Contour plot of neural models for the quadratic system.

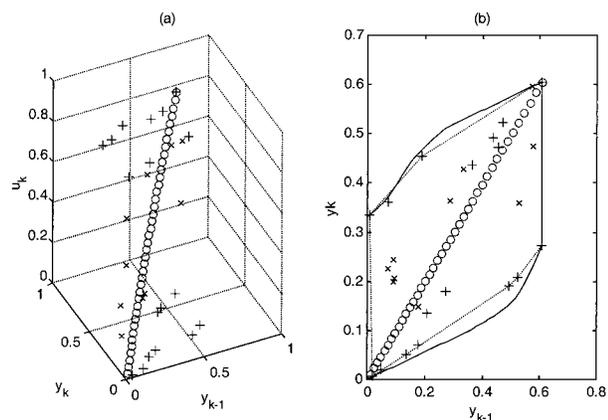


Figure 4. 3D plot of second-order dynamic system with initial PRBS data (+): partial knowledge (x), GAMIA data (O).

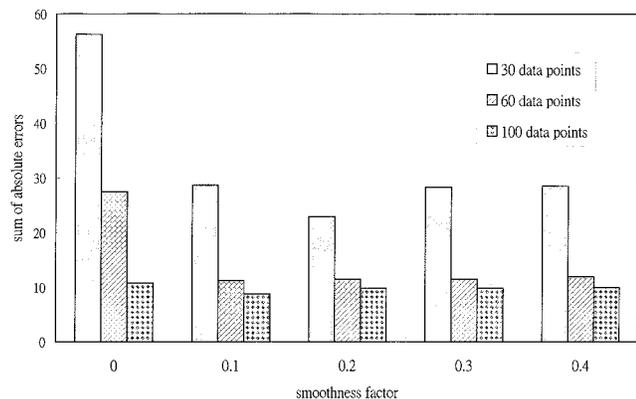


Figure 5. Effect of smoothness factor for second-order dynamic system.

Φ for this example is completely known, we uniformly sample 1093 points from Φ as the testing set. In Figure 5, the sum of absolute errors in the testing set is plotted as a bar chart versus the smoothing factor λ with the data size of Ω_0 as a parameter. The sum of absolute errors in each case without considering the smoothness factor, that is, $\lambda = 0$, is almost twice the magnitude compared to its counterpart taking the smoothness factor into consideration. It may be noted that if the initial data set is small, the effect of smoothness is significant. Smoothing the initial data can decrease errors caused by noisy data or by data extrapolation due to insufficient data. ANN modeling is a kind of curve regression. The optimal setting of the smoothness factor, λ becomes insignificant when the size of Ω_0 is large.

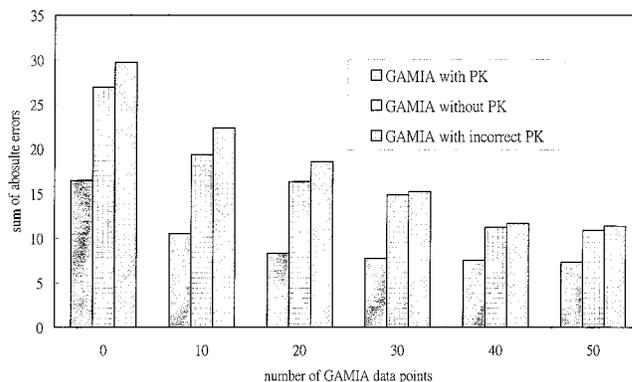


Figure 6. Effect of partial knowledge (PK) data for second-order dynamic system.

Figure 5 indicates that the optimal value of the smoothness factor, λ changes with the number of data in Ω_0 .

After the conditional AIC of eq 8 is solved, the parameters in the objective function eq 7 are then determined as $(\lambda, \beta) = (0.3488, 2.9605)$ for the case of 20 initial data. According to the GAMIA algorithm proposed in the previous section, more and more experimental data are suggested and performed during the iterative improvement of process modeling. Figure 6 demonstrates the importance of partial knowledge provided in eq 21. Given 20 initial data points from eq 20, the sum of absolute errors in the testing set decreases along with the increase of GAMIA data. The incorporation of partial knowledge data dramatically reduces the sum of errors. As shown in Figure 6, 10 GAMIA data points with partial knowledge data outperform 50 GAMIA data points without partial knowledge.

In the case of significant incorrect partial knowledge, assume that the following steady-state equation is implemented

$$0.252y_s^2 - 0.64y_s + u_s/3 = 0 \quad (23)$$

against the correct steady state eq 20. Assuming the initial 20 noise free experimental data are available and solving the conditional AIC eq 8, weighting of $(\lambda, \beta) = (0.298, 0.03768)$ is obtained. Compared with the previous case, this makes sense since the partial knowledge is incorrect. It may be noted from Figure 6 that the incorrect partial knowledge basically deteriorates the quality of the ANN model. However, the implementation of the conditional AIC reduces the effect of the incorrect partial knowledge by giving a smaller parameter for the third term in the object function eq 7.

4.3. Example 3: Multicomponent Distillation System. Figure 7 gives the schematic diagram of a multicomponent distillation system. The feed stream consists of benzene, toluene, and xylene. The design parameters of the column and the steady-state profiles are listed in Table 1. The product in the top stream is highly purified benzene. The operation objective of the system is to control the top and bottom temperatures by manipulating the reflux flow rate and the heat duty of the reboiler. The model of the distillation column is to be identified as follows

$$[y_{k+1}^1, y_{k+1}^2] = \mathbf{h}(m_k^1, m_k^2, y_k^1, y_{k-1}^1, y_k^2, y_{k-1}^2) \quad (24)$$

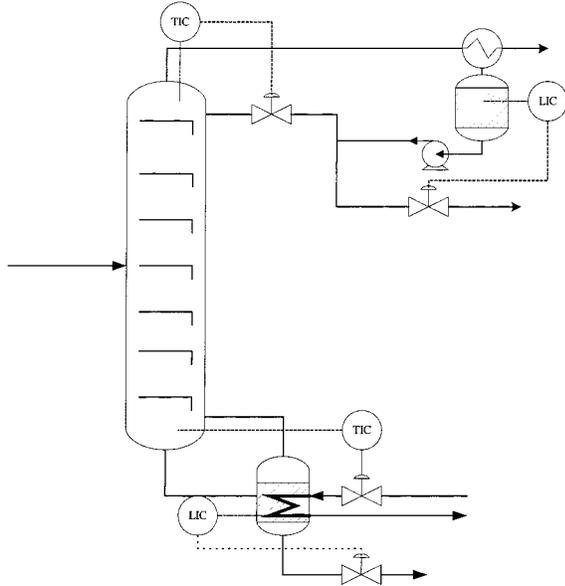


Figure 7. Schematic plot of the distillation tower.

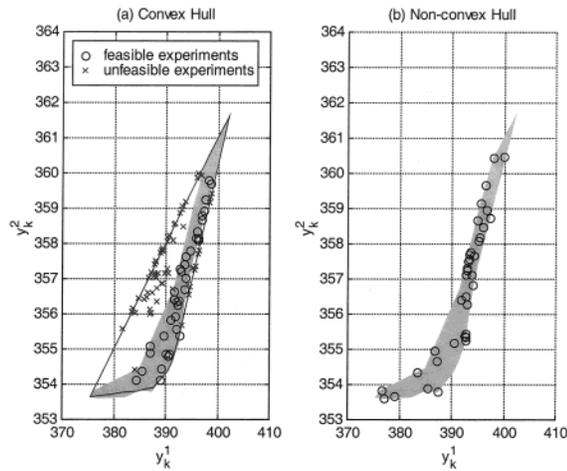


Figure 8. Suggested experimental points of the convex hull and the nonconvex hull.

Table 1. Steady-State Profile of the Distillation Column

design parameter	value	tray number	temp (K)	benzene composition (%)
top product temp (K)	389.5			
bottom product temperature (K)	354.1	1	389.5	0.028
feed flow rate (lb mol/h)	200	2	385.0	0.060
feed temperature (K)	353.2	3	381.8	0.110
feed benzene composition (%)	50	4	378.5	0.178
feed toluene composition (%)	35	5	375.1	0.260
feed xylene composition (%)	15	6	372.0	0.346
reflux flow rate (lb mol/h)	130	7	369.4	0.421
reboiler heat duty (kBTU/h)	3350	8	367.6	0.476
distillate flow rate (lb mol/h)	99.63	9	366.5	0.513
bottom product flow rate (lb mol/h)	100.37	10	365.8	0.535
operating pressure (atm)	1	11	363.2	0.599
tray number	18	12	361.4	0.656
feed tray	10	13	359.9	0.715
tray efficiency	1	14	358.4	0.776
column diameter (ft)	4.5	15	357.0	0.838
reflux drum diameter (ft)	4	16	355.8	0.893
sump diameter (ft)	4.5	17	354.8	0.940
tray area (ft ²)	15.9	18	354.1	0.976

where y_k^1 is the bottom temperature, y_k^2 is the top temperature, m_k^1 is the reboiler heat duty, and m_k^2 is the reflux flow rate.

In this case, we assume that 120 noisy dynamic plant data and a steady-state equation are available. Thirty-six noise free data derived from the steady-state equation³² represent the partial plant knowledge. The conditional AIC eq 8 is solved to find the parameters of the ANN training objective function eq 7. The GAMIA strategy is applied to determine the extended experimental data set. The extended experiments are performed for both convex and nonconvex hulls, as shown in Figure 8, where the shaded area indicates the nonconvex hull while the solid line indicates the convex hull. The first 100 experimental design points based on the convex hull scheme may suggest infeasible experiments denoted by (x) as shown in Figure 8a while the nonconvex hull scheme shown in Figure 8b would always suggest feasible experiments denoted by (o). The performances of these two ANN's are compared on the basis of the 1200 PRBS testing data set. The total absolute error of the testing set is 127 for the case of the convex assumption but only 84.4 for the nonconvex hull.

The ANN models for the column are built on the basis of the following schemes:

(1) ANN models trained by 140 GAMIA experimental data as well as 120 initial data with smoothness and partial knowledge terms included;

(2) ANN models trained by 600 operating data generated by the random amplitude sequence (RAS);³³

(3) ANN models trained by 600 operating data generated by the pseudorandom binary sequence (PRBS).³¹

The MPC controller used in servo control is implemented by the following objective function.

$$\min \sum_{j=1}^2 \sum_{i=1}^8 (\hat{y}^j(k+i|k) - y_{sp}^j) + 0.035 \sum_{i=1}^2 [\Delta m_1^i/3 + \Delta m_2^i] \quad (25)$$

where $\hat{y}(k+i|k)$ is the predicted output calculated from current time k and y_{sp}^j is the set point. All the parameters, horizons, and penalties in eq 25 are well-tuned according to our knowledge to MPC approaches.

Figure 9 compares the servo behaviors of MPC controllers based on ANN models trained in cases 1, 2, and 3. The set points of the top and bottom temperatures are changed at $t = 0$ and $t = 300$, respectively. The temperature measurements are contaminated by normal distributed noise ($\epsilon \in N(0,0.05)$). The control results in Figure 9 show that the models developed by the GAMIA approach are much superior to other cases. Adding up the IAE in the top temperature control loop and the bottom temperature control loop, the total IAE in GAMIA is only half of that in the RAS and PRBS cases.

The MPC controller used in regulation³⁴ is set by the following objective function under the assumption of the truth of the principal of superposition:

$$\min \sum_{j=1}^2 \sum_{i=1}^8 (\hat{y}^j(k+i|k) - y_{sp}^j - (\hat{y}_k^j - \hat{y}^j(k|k-1))) + 0.035 \sum_{i=1}^2 [\Delta m_1^i/3 + \Delta m_2^i] \quad (26)$$

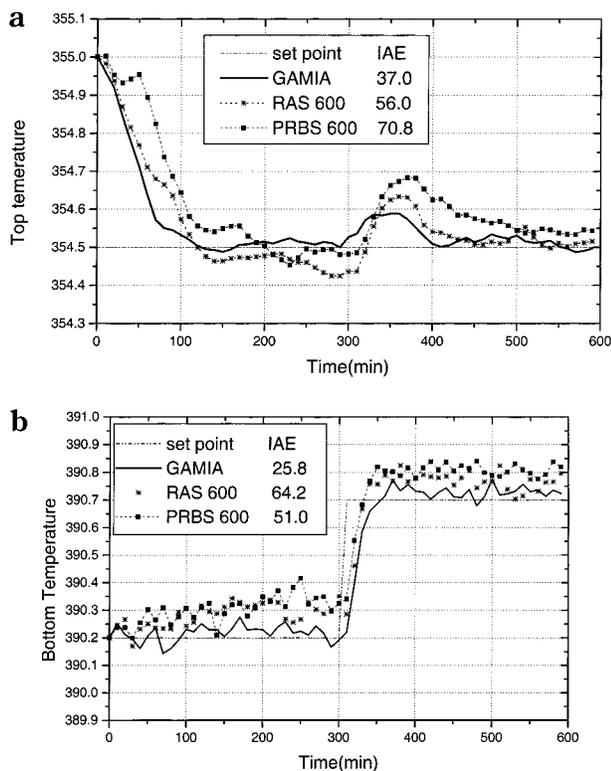


Figure 9. Servo control performance comparison: (a) top temperature control loop; (b) bottom temperature control.

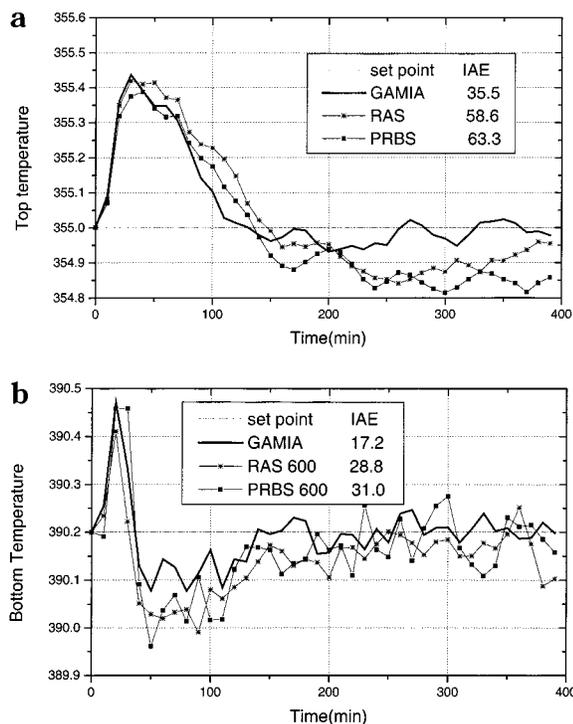


Figure 10. Regulation control performance comparison: (a) top temperature control loop; (b) bottom temperature control.

Figure 10 demonstrates the regulation behavior of MPC controllers based on ANN models trained in cases 1, 2, and 3. The composition of benzene is suddenly decreased by 5%, and that of toluene is up by 5%. Figure 10 shows that the model developed by this work is still superior to all other cases. Adding up the IAE in the two control loops, the total IAE in the GAMIA case, that is, 52.7, is much better than that in the RAS case, that is, 87.4, and that in the PRBS case, that is, 94.3.

Table 2. Regulation and Servo Control Performance Comparison

	bottom IAE(R)	top IAE(R)	total IAE(R)	bottom IAE(S)	top IAE(S)	total IAE(S)
GAMIA	17.2	35.5	52.7	25.8	37.0	62.8
RAS 1200	14.8	48.5	63.3	34.6	61.8	96.4
PRBS 1200	29.2	58.2	77.7	54.2	47.4	111.6

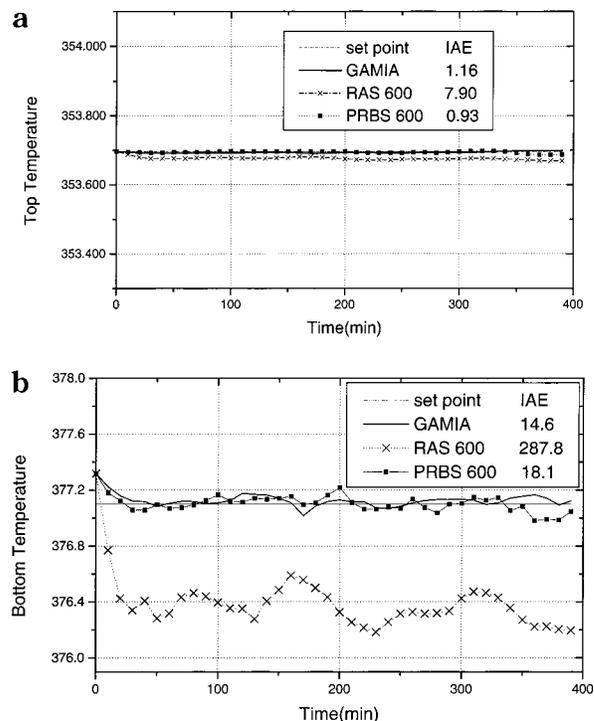


Figure 11. Regulation control performance comparison: (a) top temperature control; (b) bottom temperature control.

However, assume that RAS and PRBS may double their data size to 1200. Table 2 gives the IAE for all servo and regulation results. The GAMIA still appears superior to all other models with a much smaller training data set.

Note that, for all RAS and PRBS approaches, some unexpected extrapolation may happen, as discussed in the previous sections, for instance, if the top set point is at 353.7 K and the bottom set point is at 377.3 K. A small change in the set point of the bottom temperature to 377.1 K is implemented; then the results are shown in Figure 11; that is, the RAS ANN model fails to maintain both top and bottom at their set points.

From Figures 9–11 together with Table 2, the results clearly indicate that the model based on GAMIA works almost perfectly in all three situations while models based on RAS and PRBS perform fine in some situations and fail in other situations.

5. Conclusions

In this study, we combine initial data, partial knowledge data, along with experimental design data to train an ANN model. In the experiment design step, the constraint of a nonconvex hull guarantees the points suggested by experimental design would be in a feasible region. During modeling, noisy data are smoothed by taking the smoothness factor into consideration.

In the first example of a quadratic system, we demonstrate the effectiveness of the smoothness factor in the noisy environment. In the second example of a

dynamic system, the smoothness factor and partial knowledge are incorporated into the GAMIA algorithm. It is shown that partial knowledge, though being incorrect, may be as effective as the smoothness factor in enhancing model quality. In the final example, a realistic multiple-component distillation system is simulated. The results show that this approach is very useful for the development of gray box models that can be implemented for multivariable nonlinear model predictive control.

The new experimental points in our work are based on the minimization of the information free energy by varying the location of new data points.¹⁹ The widely accepted Taguchi method is built on an orthogonal table and maximization of the signal-to-noise ratio. The Taguchi method only suggests an individual manipulated variable operating at a low or high level. The evolutionary improvement mechanism is not well defined in the Taguchi method. Both the Taguchi method and the response surface methodology design manipulated variables only. However, we have to consider manipulated variables and response variables at once in order to design new experimental points. The traditional experimental design approach would not be appropriate in our work.

Acknowledgment

The authors thank the National Science Council, Taiwan, ROC, for financial support of this work through Grants NSC88-2214-E007-009 and NSC88-CPC-E007-015.

Nomenclature

C = convex hull
 NC = nonconvex hull
 \mathbf{x} = state variables
 \mathbf{m} = system inputs
 \mathbf{y} = system outputs
 T = sampling time
 \mathbf{h} = discrete time, input–output model
 \mathbf{h}' = discrete time, input–output model indicated by partial knowledge
 h_i = output of the i th neuron of the hidden layer in neural networks
 $\bar{\mathbf{y}}_k$ = on-line measurements contaminated by white noise
 $\hat{\mathbf{y}}_{k+1}$ = dynamic system output indicated by partial knowledge
 $\hat{\mathbf{y}}_{k+1}^1$ = dynamic system output predicted by neural networks
 w^h = weightings between the input layer and the hidden layer in the neural network
 w^o = weightings between hidden layer and the output layer in the neural network
 \mathbf{w} = weightings in the neural network model, $\{w^h, b^h, w^o, b^o\}$
 $J(\mathbf{w})$ = training objective function of neural networks
 b^o = bias of nodes in the output layer
 b^h = bias of nodes in the hidden layer
 N_G = number of grid points where the smoothness factor is calculated
 N = number of experimental data points
 K = number of training data points
 S = information entropy
 T = information temperature
 H = information enthalpy
 G = information free energy
 ΔG_{STOP} = stopping criteria of the free energy change

E_{stop} = stopping criteria for the difference between \mathbf{X}_{pre} and \mathbf{X}_{exp}
 R = number of partial knowledge points
 Q = number of events in Φ''
 p = probability measure
 u = dummy variable
 c = constant used in information temperature scheduling
 z = constant used in information temperature scheduling
 $\hat{\mathbf{y}}(k + 1|k)$ = neural model predicted output calculated from current time k
 \mathbf{y}_{sp} = set point
 b = support function of convex hull
 \mathbf{u} = unit vector
 $D(r)$ = support function of convex hull in a polar coordinate system
 M = a stationary process
 $diam$ = diameter of a set
 a_n = a constant sequence
 L = a sequence of process

Greek Symbols

λ = penalty of smoothness factor
 β = penalty of partial knowledge
 Φ = feasible region
 Φ' = feasible region contaminated by white noise
 Φ'' = discretized approximation of feasible region
 σ^2 = variance of normal probability distribution
 μ = fuzzy membership function of a possible event
 ϕ^i = element of feasible region Φ
 Ω = experimental data set
 Ω' = experimental data set contaminated by white noise
 ω = dynamic event set
 ω_i = element of experiment data set
 Ψ = partial knowledge
 ψ^i = element of partial knowledge
 χ = input vector of neural networks
 $\zeta = (\lambda, \beta)$ = penalties to be determined in conditional AIC
 Λ = Gaussian normal process (white noise)
 κ = element of Λ

Subscripts

k = current time
 s = input order of the dynamic system \mathbf{h}
 s' = output order of the dynamic system \mathbf{h}
 $m1, m2, m3, m4, m5, m6$ = input order of the dynamic system
 $n1, n2, n3, n4, n5, n6$ = output order of the dynamic system

Superscripts

1 = first system output
 2 = second system output

Special Operators

\oplus = Minkowski addition of sets
 \underline{w} = weak convergence

Appendix

1. Lemma (Minkowski Addition of Two Compact Sets). Denote Minkowski addition \oplus for two sets $\Lambda = \{d_i | i = 1, \dots\}$ and $\Omega = \{\omega_j | j = 1, \dots\}$

$$\Lambda \oplus \Omega = \{d_i + \omega_j | \forall d_i \in \Lambda, \omega_j \in \Omega, i = 1, \dots, j = 1, \dots\}$$

If Λ and Ω are compact, then $\Lambda \oplus \Omega$ is compact.

Proof: Define the diameter of a set as

$$diam(W) = \sup_{\omega_i, \omega_j \in \Omega} |\omega_i - \omega_j|$$

Both $diam(\Omega)$ and $diam(\Lambda)$ are finite because Ω and Λ

are compact.

$$\begin{aligned} \text{diam}(\Omega \oplus \Lambda) &= \sup_{x_i, x_j \in \Omega \oplus \Lambda} |x_i - x_j| \leq \sup_{\substack{\omega_i, \omega_j \in \Omega \\ d_p, d_q \in \Lambda}} \omega_i + \\ & d_p - \kappa_j - \omega_q \leq \sup_{d_i, d_j \in \Lambda} |\kappa_i - \kappa_j| + \sup_{\omega_p, \omega_q \in \Omega} |\omega_p - \omega_q| \\ & d_q = \text{diam}(\Lambda) + \text{diam}(\Omega) \end{aligned}$$

Therefore, $\text{diam}(\Omega \oplus \Lambda)$ is finite. And $\Lambda \oplus \Omega$ is compact.

2. Proof of Proposition 2-1. (a) Convergence of Convex Hull for a Gaussian Normal Process. Λ can be viewed as a stochastic process. Let K_d be a collection of a nonempty compact convex subset in R^d . Define the support function of $K \in K_d$ as

$$b_k(u) = \sup_{k \in K} \{ \mathbf{k} \cdot \mathbf{u} \mid \mathbf{u} = 1 \}$$

where \cdot is the inner product in R^d and $\mathbf{k} \in K$.

The support function b_k is a continuous function on the unit ball in R^d . Because $\Lambda = \{k_1, \dots, k_N\}$ is independent and identically distributed in R^2 , the support function of the convex hull $C(\Lambda)$ in the polar coordinate is

$$\sup_i \{ b_{x_i}(r) \} = \sup_x \{ B_i(r) \} = D(r)$$

where $\{B_i\}$ is a sequence of gaussian normal processes on $[0, 2\pi]$ and r is the phase angle. The area of the convex hull is

$$\frac{1}{2} \int_0^{2\pi} \left[D^2(r) - \left(\frac{\partial D(r)}{\partial r} \right)^2 \right] dr$$

And the perimeter of the convex hull is

$$\int_0^{2\pi} D(r) dr$$

A direction computation of the above equation would be difficult since $D(r)$ does not have a simple form. Define a new sequence of processes as

$$L_{n,i}(t) = a_n [B_i(a_n^{-1}t) - a_n], t \in [0, T], T < \infty$$

$\{a_n\}$ is defined as³⁵

$$a_n = (2 \log n^2) - \frac{1}{2} (\log \log n + \log 4\pi) / (2 \log n)^{1/2}$$

then

$$\lim_{n \rightarrow \infty} \sup_i |L_{n,i}(t)| \xrightarrow{w} M$$

where M is a stationary process and \xrightarrow{w} stands for weak convergence.³⁶ We can conclude that $\{B_i\}$ is limited and that $D(r)$ is finite and some properties of the convex hull of Λ , such as area and perimeter converge. Therefore, $C(\Lambda)$ and Λ are compact.

(b) Convergence of Convex Hull for Noisy Data. Assuming that $C(\Omega)$ and Ω are compact, since the dynamic event of a deterministic process is limited to the available control input resource, $C(\Lambda)$ and Λ are compact by proposition 2.1(a).

Given $\Omega' = \{x_i = d_i + \kappa_i \mid i = 1, \dots, N, \kappa_i \in \Lambda, d_i \in \Omega\}$, then $\Omega' \subset \Lambda \oplus \Omega$. According to lemma 2.1, $C(\Lambda \oplus \Omega)$ and $\Lambda \oplus \Omega$ are compact. Then $C(\Omega')$ is compact too.

3. Proof of Proposition 2-2 (Convergence of Nonconvex Hull)

$$N(\Omega') = \bigcup_{i=1}^m C(\Omega'_i)$$

According to proposition 2-1, $C(\Omega'_i)$ converges and $\text{area}(C(\Omega'_i))$ is finite.

$$\text{area}(N(\Omega')) = \bigcup_{i=1}^m \text{area}(C(\Omega'_i)) \leq \sum_{i=1}^m \text{area}(C(\Omega'_i))$$

$\text{area}(N(\Omega'))$ is finite, and $N(\Omega')$ converges.

Literature Cited

- (1) Garica, C. E.; Morari, M. Internal model control 1. a unifying review and some new results. *Ind. Eng. Chem. Res.* **1982**, *21*, 308–323.
- (2) Culter, C. R.; Ramaker, B. L. Dynamic matrix control, a computer control algorithm. Presented at the Eighty-Sixth Meeting of the AIChE, Houston, April 1979.
- (3) Savkovicstevanovic, J. Neural net controller by inverse modeling for a distillation plant. *Comput. Chem. Eng.* **1996**, *20*, 935–930.
- (4) Su, H. T.; McAvoy, T. J. Long term prediction use recurrent neural networks—A parallel training approach. *Ind. Eng. Chem. Res.* **1993**, *31*, 1338–1352.
- (5) Sentoni, G. Approximate Models for Nonlinear Process Control. *AIChE J.* **1996**, *42*, 2240–2250.
- (6) Thompson, M. L.; Kramer, M. A. Modeling Chemical Processes Using Prior Knowledge and Neural Networks. *AIChE J.* **1994**, *40*, (8), 1328–1340.
- (7) Psychogios, D. C.; Ungar, L. H. Direct and Indirect model based control using artificial neural networks. *Ind. Eng. Chem. Res.* **1991**, *30*, 2564.
- (8) Tsen, A. Y.; Jang, S. S. Predictive control of quality in batch polymerization using hybrid ann models. *AIChE J.* **1996**, *42*, 455–464.
- (9) Van Can, H. J. L.; Luyben, K. C. A. M.; Heinen, J. J. Strategy for Dynamic Process Modeling Based on Neural Networks in Macroscopic Balances. *AIChE J.* **1996**, *42*, 3403–3418.
- (10) Vancan, H. J. L.; Tebraake, H. A. B.; Dubbelman, S.; Hellinga, C.; Luyben, K. C. A. M.; Heijnen, J. J. Understanding and applying the extrapolation properties of serial gray-box models. *AIChE J.* **1998**, *44*, 1071–1089.
- (11) Akaike, H. Fitting auto-regressive models for prediction. *Ann. Inst. Stat. Math.* **1969**, *21*, 243–247.
- (12) Akaike, H. Statistical predictor identification. *Ann. Inst. Stat. Math.* **1970**, *22*, 203–217.
- (13) Stoica, P.; Janssen, P.; Soderstrom, T.; Eykhoff, P. Model-structure selection by cross-validation. *Int. J. Control* **1985**, *43*, 1941–1878.
- (14) Tong, H. *Nonlinear time series: a dynamical system approach*; Oxford University Press: New York, 1990.
- (15) Dong, D.; McAvoy, T. J. Nonlinear Principal component analysis—based on principal curve and neural networks. *Comput. Chem. Eng.* **1990**, *20*, 65–78.
- (16) Johansen, T. A. Identification of Nonlinear Systems using Empirical Data and Prior Knowledge—An Optimization Approach. *Automatica* **1994**, *32*, 337–356.
- (17) Raju, G. K.; Cooney, C. L. Active learning from process data. *AIChE J.* **1998**, *44*, 2199–2211.
- (18) 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.
- (19) Lin, J. S.; Jang, S. S. Nonlinear dynamic artificial neural network modeling using an information theory based experiment design approach. *Ind. Eng. Chem. Res.* **1998**, *37*, 3640–3651.

- (20) Cavendish; Field, D. A.; Frey, W. H. Automatic triangulation of arbitrary planar domains for the finite element method. *Int. J. Numer. Methods Eng.* **1974**, *8*, 679–686.
- (21) Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423.
- (22) Shieh, S. S. Automatic knowledge acquisition from routine data for process and quality control. Doctoral dissertation, Washington University, St. Louis, 1992.
- (23) Peace, G. L. *Taguchi methods: a hands-on approach*; Addison-Wesley: Reading, MA, 1993.
- (24) Baratti, R.; Bertuccio, A.; Darold, A.; Morbidelli, M. A composition estimator for multicomponent distillation columns development and experiment test on ternary mixtures. *Chem. Eng. Sci.* **1998**, *53*, 3601–3612.
- (25) Doyle, F. J., III; Shaw, M. A. Multivariable nonlinear control application for a high purity distillation column using a recurrent dynamic neuron model. *J. Process. Control* **1996**, *7*, 255–268.
- (26) 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.
- (27) Rhodes, C.; Morari, M. The false nearest neighbors algorithm: An Overview, *Comput. Chem. Eng.* **1997**, *21*, S1149–1154.
- (28) Doherty, S. K.; Gomm, J. B.; Williams, D. Experimental design considerations for nonlinear system identification using neural networks. *Comput. Chem. Eng.* **1997**, *21*, 327–346.
- (29) 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), 2005–203.
- (30) Blahut, R. E. *Principles and Practice of Information Theory*; Addison-Wesley: Reading, MA, 1991.
- (31) Goodwin, G. C.; Payne, R. L. *Dynamic system identification: experiment design and data analysis*, 1977.
- (32) Holland, C. D.; Liapis, A. I. *Computer Methods for Solving Dynamic Separation Problems*; Prentice Hall: New York, 1979.
- (33) Pottmann, M.; Seborg, D. E. A radial basis function control strategy and its application to a pH neutralization process. Paper Presented at Second European Control Conference, EEC'93, Groningen, The Netherlands, 1993.
- (34) Economou, G. E.; Morari, M.; Palsson, B. O. Internal Model Control. 5. Extension to Nonlinear Systems. *Ind. Eng. Chem. Process Des. Dev.* **1986**, *25*, 403–411.
- (35) David, H. A. *Order Statistics*; Wiley: New York, 1970.
- (36) Eddy, W. F. The distribution of the convex hull of a Gaussian sample. *J. Appl. Prob.* **1980**, *17*, 686–695.

Received for review May 4, 1999

Revised manuscript received August 26, 1999

Accepted September 29, 1999

IE990312E