The Enhancement of Empirical Model Capability and Optimal/Robust Design of Intractable Processes

Jun-Shien Lin, Shi-Shang Jang,* Shyh-Jye Chien, and Chen-Chi Ma

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

Shyan-Shu Shieh

Department of Occupational Safety and Hygiene, Chang Jung University, Tainan County, Taiwan

Most industrial systems are developed and improved more from experimental data than from theoretical analysis. In this work, the robust design and optimal design of products and processes are developed using an information-index-based artificial neural network response surface. The structure and training policy of such an artificial neural network model are determined by the cross-validation information index (CVI for short) developed in this work. In the case of noisy and limited experimental data, this information index is particularly useful for finding the number of nodes in the hidden layer and the weighting of the smoothness factor. The polymer composite pultrusion process is studied. Simulation and experimental results show that this novel approach is highly effective and promising.

Introduction

Product quality and stability are very important for industrial processes. In most cases, it is very expensive to derive a first principles model in process industries. Product design and quality improvements for most systems rely highly on experimental data and empirical models. In the past decade, the artificial neural network (ANN) approach has become widely used as a tool in building empirical models.[1,2] Previously, the authors[1] extended the information theory concept to derive an optimization approach for finding improved conditions that maximize the product quality or other economic objective function. Meanwhile, the ANN model can also be steadily improved by using augmented training data sets. However, the following problems remain to be addressed: (1) The complexities of artificial neural network models, i.e., the number of hidden layer nodes, should depend more on the systems themselves than on the quantity of experimental data. Few discussions on this issue have been made. (2) Artificial neural networks have been successfully demonstrated in modeling process systems. However, they fail when the experimental data are noisy and expensive to obtain.

The objective of this work is to tackle the above problems. Unlike our previous work,[1] this work focuses on modeling for the cases of scarce and noisy data. The current study also derives a novel index for solving this problem. Some theoretical analysis of this index is provided.

Artificial neural networks (ANNs) have been applied to solve various problems in the chemical engineering domain. Many researchers have concentrated on model construction and used these models to predict process behavior.[3−5] Furthermore, some have derived dynamic ANN models for the purpose of process control.[6−9] Partial-knowledge-based ANN control is also under thorough study.[10−14] However, all of the above works did not consider the appropriateness of the size of the neural network models with or without prior knowledge. On the other hand, the determination of ANN model structures has been broadly studied in the field of statistics. Akaike's information criteria (AIC) is widely implemented in determining the number of hidden layer nodes.[15,16] However, the approach is difficult when training data are very noisy. Conditional AIC[17] is applicable in some cases but provides unsatisfactory results in the case of limited and noisy data.

During product development stage, experimental data might not only be noisy, but might also be quite limited and expensive to obtain. Stoica[18] proposed the use of cross-validation to find model parameters in this situation. Galatsanos and Katsaggelos[19] extended the approach to estimate the smoothness parameter, which is also called the regularizing parameter in ANN modeling.

The basic theme in increasing the generalization capability of neural network model can be categorized into the following three issues: (1) Model selection.[15] In this work, we consider model selection as finding the optimal number of hidden nodes. However, in some other studies, model selection includes neuron-pruning.[20] (2) Regularization approach.[21] A smoothness term can be added to the neural training objective function. This term is often a constraint on the model behavior or penalty on weighting. (3) Early stopping.[22] A common phenomenon of neural training is that training errors decrease monotonically as training increases, but generalization errors, reflecting the real error between the model and system, reach a minimum and then increase.

More specifically, this study aims to incorporate the following considerations into an ANN model for optimal/robust design of a system with highly noisy and few available data: (1) a reasonable number of hidden layer nodes in an ANN model and (2) a penalized term added to the training objective function of the ANN model to reflect the fact that the response surface of a real system is smooth.

In this work, we derive a novel information index to determine the ANN model structure and smoothness

* Corresponding author. E-mail: ssjang@che.nthu.edu.tw.

10.1021/ie000543h CCC: $20.00
Published on Web 08/04/2001

parameter simultaneously. Simulations and experimental studies for a polymer composite pultrusion process are implemented to verify the proposed approach.

Pultrusion is one of the most popular processes for polymer composite manufacturing. Mathematical models for temperature and monomer concentration profiles are well-established.[23−25] As with many physical models, these models are based on several ideal assumptions. Furthermore, only models for product physical properties are of industrial interest. Some correlations for resin and product physical properties are available.[24,26]

Armed with the above-mentioned models, we still have difficulties in determining the optimal design for a pultrusion process. In particular, the objective of the experiment is to seek the optimal operating conditions for maximizing the fracture strength, tensile strength, impact strength and production capacity. During experimentation, we experienced high noise levels in mechanical property measurements, even within the same batch of production. The *S*/*N* ratio is extremely poor and can be as low as 1. In addition, it takes 3 h to run a batch of production and even longer to measure the physical properties. In the considerable period, the experimental data are still limited. With scarce and noisy data, it is hard to obtain a good empirical model, not to mention the optimal design. Therefore, we consider the pultrusion process to be intractable, because it is not easy to model by first principles or by any empirical method. According to our knowledge, this is a very common phenomenon in the product-development phase for industry. The challenges provided by the pultrusion process motivate us to propose a novel response surface methodology. Once a good empirical model is obtained, optimal operating conditions can be acquired using the previously proposed optimization approach.[1]

Experimental design for product quality improvement is traditionally based on a polynomial response surface.[27] The authors[1] proposed the implementation of an ANN response surface. Our previous study showed that system noises could only be reconciled by model training when they are small. However, this might not be the case for many industrial systems. In high-noise systems, the Taguchi method[28] is widely used in the industry to locate the operating conditions where the signal-to-noise ratio (*S*/*N*) is maximized. However, implementation of the Taguchi approach can only determine the robust conditions at the predetermined levels. In this work, CVI-based ANN response surfaces are constructed to obtain the robust operating conditions in an evolutionary manner with precise values. The major contributions of this work can be itemized as follows: (1) A novel information index is derived to determine the structure and parameters of an empirical model, such as an ANN, in the case that data are scarce and noisy. (2) The concept of robust design, maximization of *S*/*N* ratio, is experimentally solved using a response surface for the first time. (3) The process design and robust design problem of an intractable process, namely, the pultrusion process is solved using the proposed modeling technique.

This paper is organized as follows. In the next section, we derive a CVI-based ANN model whose structure and smoothness factor are determined using a novel information index that combines the concepts of cross-validation and Akaike's information criteria. In section 3, we discuss the distinction between robust design and optimal design. Robust design means having a stable product quality by finding the optimal operating conditions that maximize the *S*/*N* ratio. On the other hand, optimal product design is to find the operating condition, which maximizes production objective function. In the same section, robust design and optimal design problems are formulated and solved using the desired response surface and iterative experimental data. The mathematical model and experimental setup for the pultrusion system are described in section 4. In section 5, three examples are presented to demonstrate the application of the proposed methodology. The first example, curve fitting of a quadratic function, shows the usefulness of the proposed index in modeling a system with limited and noisy data. The second and third examples are the simulation and real experiment of a pultrusion process that is considered to be extremely intractable. The second example presents comparisons of the proposed index and existing indices in determining the model structure and the resulting model capability. Then, we apply the proposed index to model a real pultrusion process. The acquired ANN model is subsequently used to find the optimal operating conditions. Also robust design is conducted as the data are very noisy. Finally, some concluding remarks are given in section 6.

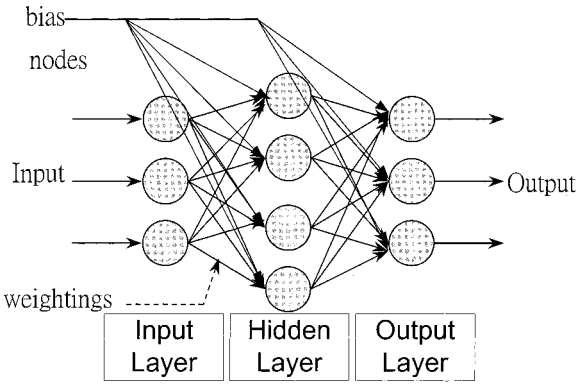

Figure 1. Schematic plot of a single hidden layer neural network model.

2. Development of the CVI-Based ANN Response Surface

Consider a process system S that generates experimental data with an input vector $\mathbf{x} \in R^N$, an unknown disturbance vector $\mathbf{z} \in R^P$, and a system output (physical properties of the product that can be measured) vector $\mathbf{y} \in R^M$, such that

$$\mathbf{y} = S(\mathbf{x},\mathbf{z}) \tag{1}$$

Assume that a set of experimental data $\Omega = \{(\mathbf{x}_i,\bar{\mathbf{y}}_i)|i = 1, ..., Q\}$is available, where $\bar{\mathbf{y}}$ denotes the vector of the measured physical properties. We denote a response surface model for system S by

$$\hat{\mathbf{y}} = \mathrm{RS}(\Xi,\mathbf{w},\mathbf{x}) \tag{2}$$

where $\Xi \in R^L$ represents parameters of the model and $\mathbf{w} \in R^K$ is a weighting vector for the nonlinear functions in the model, such as the neurons' weighting for an ANN model.

2−1. Theoretical Background. Consider an ANN model as shown in Figure 1. Assume that *N* input nodes $x = (x_1,x_2,...,x_N)$ and a hidden layer with *H* nodes are

implemented; then the output signal from the hidden layer with respect to an input signal, x_j, can be derived as

$$h_i(x) = \tanh\left(\sum_{j=1}^{N} \mathbf{w}_{ij}^{\,h} x_j + \mathbf{b}_i^{\,h}\right)$$

$$\mathbf{w}^h \in R^H \times R^N,\ \mathbf{b}^h \in R^H \tag{3}$$

if one implements a hypertangent transfer function. Herein, for simplicity, we only implement a single hidden layer, but the results in this work can easily be extended to multilayer ANN systems. It can be easily shown that the mth output of the model is given by

$$\hat{y}^m = \sum_{i=1}^{H} w_{mi}^{o} h_i(\mathbf{x}) + b_m^{o}$$

$$\mathbf{w}^o \in R^M \times R^H,\ \mathbf{b}^o \in R^M \tag{4}$$

We denote the following weightings of the model in eq 2 for an ANN model

$$\mathbf{w} = \{\mathbf{w}^h, \mathbf{b}^h, \mathbf{w}^o, \mathbf{b}^o\}$$

Without the parameter set Ξ in eq 2, the common training strategy for an ANN model in eqs 3 and 4 is to solve the optimization problem

$$\min_{\mathbf{w}} \sum_{i=1}^{Q} \|\hat{\mathbf{y}}_i - \bar{\mathbf{y}}_i\|^2 \tag{5}$$

However, the above ANN model is usually trained by a large training set such that the data noises can be filtered out by eq 5. For many processes, especially those in the product-development phase, the experimental data are noisy and scarce. To take smoothness into account, the objective function in eq 5 can be modified as

$$J(\mathbf{w}) = \sum_{i=1}^{Q} |\hat{\mathbf{y}}(x_i, \mathbf{w}) - \bar{\mathbf{y}}(x_i)|^2 + \lambda \sum_{i=1}^{N_G} \left\| \frac{\partial^2 \hat{\mathbf{y}}(x_i)}{\partial x_i^{\,2}} \right\|^2 \tag{6}$$

where N_G is the number of selected points in the response surface. Note that the last term in eq 6 can be analytically obtained from an ANN model, i.e.

$$\frac{\partial^2 \hat{\mathbf{y}}(\mathbf{x})}{\partial (\mathbf{x})^2} = \sum_{i=1}^{m} \mathbf{w}^{hT} \frac{-2}{\operatorname{diag}[\operatorname{csc} h(\mathbf{w}^h \mathbf{x} + \mathbf{b}^h)]^2} \times \operatorname{diag}[\tanh(\mathbf{w}^h \mathbf{x} + \mathbf{b}^h)] \operatorname{diag}[\mathbf{w}_i^o]\, \mathbf{w}^h \tag{7}$$

where $\mathbf{w}_i^o$ is the ith row of $\mathbf{w}^o$.

The determination of the parameter set $\Xi = \{H, \lambda\}$, where H is the number of hidden nodes, is one of objectives of this study.

Definition 2−1. Kullback−Leibler Information Entropy. Given a training set $\Omega = \{(\mathbf{x}_i, \bar{\mathbf{y}}_i) | i = 1, ..., Q\}$ and a system S with a model RS as in eqs 1 and 2, respectively, assume that all of the elements in Ω are independent; then the Kullback−Leibler information entropy is denoted by

$$I(\mathrm{S}, \mathrm{RS}) = \int \mathrm{S}(\mathbf{x}) \ln \frac{\mathrm{RS}(\mathbf{x})}{S(\mathbf{x})} \mathrm{d}\mathbf{x} \tag{8}$$

Note that $I = 0$ if S = RS according to the above definition.

Proposition 2−1. Akaike's Information Criterion for ANN Models. The optimal structure for S that minimizes the Kullback−Leibler information entropy is equivalent to minimizing

$$\mathrm{AIC} = -2 \ln(\text{maximized likelihood function}) + 2(\text{number of independently adjusted parameters})$$

In the case of the ANN models in eqs 3 and 4

$$\min_{\Xi} \mathrm{AIC}(\Xi) = \ln\left[\frac{1}{Q}\sum_{i=1}^{Q} \|\hat{\mathbf{y}}(x_i, \Xi) - \bar{\mathbf{y}}(x_i)\|^2\right] + \frac{2}{Q}[\dim[\Xi, \mathbf{w}] - 1] \tag{9}$$

where Ξ is a set of parameters that is independent of the weightings $\mathbf{w}$ among the neurons.

For the proof of proposition 2−1, see Tong.[17]

The advantage in AIC is that the dimensionality of the model can be viewed as being comparable in order of magnitude to the training error. However, as the size of the training set is small, it is difficult to solve eq 6, whose objective is to determine model structure and smoothness parameters simultaneously.[29] As in all optimization problems, the two terms of eq 9 compete with each other, and AIC is a compromise. In the case of highly noisy and limited data, the solution of the above equation would prefer a larger neural network to minimize the first term. This is because the first term in AIC could be greatly reduced by a larger neural network (an increase in the second term in AIC), and this would result in overfitting as indicated in the simulation example of the pultrusion process. Moody[20] proposed to implement cross-validation to ease this problem.

The basic idea of using cross-validation is first to separate the data set Ω into D independent subsets

$$\Omega_1, \Omega_2, ..., \Omega_D, \text{ such that } \Omega = \Omega_1 \cup ... \cup \Omega_D \text{ and } \forall\, i,j \in \{1, ..., D\},\ i \neq j,\ \Omega_i \cap \Omega_j = 0$$

Use one of the subsets as the testing set and the rest as the training set. Rotate the choice of the testing set to find min CV. Define the following cross-validation index (CV); then the model parameters in eq 2 are obtained by minimizing the index.

$$\min_{\Xi} \mathrm{CV}(\Xi) = \sum_{i=1} \sum_{x \in \Omega_i} \|\hat{\mathbf{y}}^i(\mathbf{x}, \Xi) - \bar{\mathbf{y}}^i(\mathbf{x})\|^2 \tag{10}$$

where $\hat{\mathbf{y}}^i(\mathbf{x}_j, \Xi)$ is the prediction of neural model using $\Omega - \Omega_i$ as the training data set. However, this approach is inappropriate when data are noisy and limited.[30] The problem is that size of the neural network needs to be determined implicitly via the very small number validation data. This would lead to the same problem as AIC. Nevertheless, the smoothness factor in eq 6 could be properly estimated because of the nature of cross-validation. Judging from the advantages and disadvantages of AIC and CV, a novel information index, the cross-validation information index, is derived. The definition of the new index and the proof of its conver-

gence are shown in the next section. Subsequently, the ANN response surface of a pultrusion process is determined using the new information index.

2−2. Cross-Validation Information Index (CVI). Given the conditional likelihood function for subset Ω_i, i.e.

$$L_i(\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_{Q_i}) = \left(\frac{1}{2\pi\sigma^2}\right)^{-Q_i/2} \exp\left[-\frac{1}{2\pi\sigma^2}\sum_{i=1}^{Q_i}(\hat{\mathbf{y}}(\mathbf{x}_i) - \bar{\mathbf{y}}(\mathbf{x}_i))^2\right] \quad (11)$$

where Q_i is the number of elements in Ω_i. The total conditional likelihood function for Ω is obtained as

$$L = \prod_{i=1}^{D} L_i = \left(\frac{1}{2\pi\sigma^2}\right)^{-Q/2} \exp\left[-\frac{1}{2\pi\sigma^2}\sum_{i=1}^{D}\sum_{x_i\in\Omega_i}(y(x_i) - \hat{y}(x_i))^2\right] \quad (12)$$

Proposition 2−2. Cross-Validation Information Index. Assume that the conditional likelihood function based on cross-validation in eq 12 is maximized; then the minimization of Kullback−Leibler information entropy is equivalent to minimization of the following index

$$\min_{\Xi} \mathrm{CVI}(\Xi) = \ln\left(\frac{1}{Q}\sum_{i=1}^{D}\sum_{x\in\Omega_i}\|\hat{\mathbf{y}}^i(x,\Xi) - \bar{\mathbf{y}}^i(x)\|^2\right) + \frac{2}{Q}[\dim(\Xi,\mathbf{w}) - 1] \quad (13)$$

For the proof of proposition 2−2, see the Appendix.

The above index is termed the cross-validation information index (CVI) in this work.

Proposition 2−3. Unbiased Estimation. Given a set of experimental data $\Omega = \{(\mathbf{x}_i,\mathbf{y}_i)|i = 1, ..., Q\}$, assume that all data are independent observations or Markov-dependent observations. Denote Ξ^* as the optimal settings for $Q \to \infty$ and $\hat{\Xi}$ as the solution suggested by CVI when Q data are available; then

$$\hat{\Xi} - \Xi^* \approx \frac{1}{\sqrt{Q}}N(0,\mathbf{M}) \quad (14)$$

where $\mathbf{M}$ is a constant matrix.

For the proof of proposition 2−3, see the Appendix.

According to the above proposition, the solution approaches the optimum when the amount of data in the training set becomes larger and larger. When the data set is large enough, the solutions obtained by CVI, AIC, and CV should be the same.[18,31] However, the superiority of the CVI is the inclusion of a cross-validation concept into the AIC.

Proposition 2−4. The Effect of Cross-Validation. Given a set of experimental data $\Omega = \{(\mathbf{x}_i,\mathbf{y}_i)|i = 1, ..., Q\}$, denote $\hat{\Xi}_D$ as the solution suggested by CVI when Q data are available and divided into D subsets for CVI computation; then

$$|\hat{\Xi}_D - \Xi^*| = O\left(\frac{1}{\sqrt{D}\sqrt{Q}}\right) \quad \text{as } Q \to \infty \quad (15)$$

For the proof of proposition 2−4, see the Appendix.

The above proposition suggests that dividing a training set into more training subsets makes convergence to the optimum faster. In other words, large D in eq 13 is desirable. In reality, however, with a given training set Ω, D, and λ, the evaluation of a CVI needs to train D times of sub-ANN models. Therefore, too large a value of D might become infeasible for finding an optimal setting for Ξ.

It should be noted that the new information index in eq 13 is a combination of AIC in eq 9 and CV in eq 10. CVI increases the number of hidden nodes much more conservatively than CV and AIC as the CVI considers cross-validation error and model degrees of freedom simultaneously. The overfitting problem can be avoided.

3. Robust Design and Optimal Design Using the CVI-Based ANN Model

3−1. Robust Design Using CVI-Based Model. Taguchi[32] proposes that robust design of a product involves finding operating conditions for a process such that the ratio of the product quality index q^2 (or signal, S) to the quality variance σ^2 (or noise, N) is maximized. However, the Taguchi approach can only determine the optimum at predetermined levels. Therefore, very often, it only reaches the local optimum. This work implements the response surface obtained in the last section to find the global optimum of the robust design in a feasible region because the suggested operating levels are not limited in a predetermined way.

Consider the process system in eq 1 and a training set $\Omega = \{(\mathbf{x}_i,\mathbf{y}_i)|i = 1, ..., Q\}$, where $\mathbf{x}_i$ is the design variable vector in experimental design problems and $\mathbf{y}_i$ is the measured physical property vector. To understand the behavior of the plant noises, each experimental observation ($\mathbf{x}_i$, $\mathbf{y}_i$) is to be repeated R times. Therefore, the mean, $E\mathbf{y}_i$, and variance, σ^2, for each product property, $\mathbf{y}_i$, can be estimated, where E is the expectation operator. In many cases, it is desirable to define the following product quality index as a function of the mean of the product properties

$$q = f(E\bar{\mathbf{y}}) \quad (16)$$

where q is a scalar. Substituting q_i for $\mathbf{y}_i$, we obtain

$$\Omega = \{(\mathbf{x}_i,q_i)|i = 1, ..., Q\}$$

With the above training set, we can obtain a product quality response surface function as proposed in the previous section, i.e., the following CVI-based ANN model can be obtained by solving eqs 6 and 13 iteratively

$$\hat{q} = \mathrm{ANN}_q(\Xi,\mathbf{w},\mathbf{x}) \quad (17)$$

With the same procedure, the signal-to-noise ratio response (S/N ratio) surface

$$\hat{q}^2/\sigma_q^2 = \mathrm{ANN}_{q^2/\sigma_q^2}(\Xi,\mathbf{w},\mathbf{x}) \quad (18)$$

can also be obtained. For robust design,[27] the signal-to-noise ratio (S/N) should be maximized as follows

$$\max_{\mathbf{x}} \log_{10} \hat{q}^2/\sigma_q^2$$
$$\text{s.t.} \quad \hat{q}^2/\sigma_{q^2} = \mathrm{ANN}_{q^2/\sigma_q^2}(\Xi,\mathbf{w},\mathbf{x}) \quad (19)$$

It should be noted that the solution of optimization for

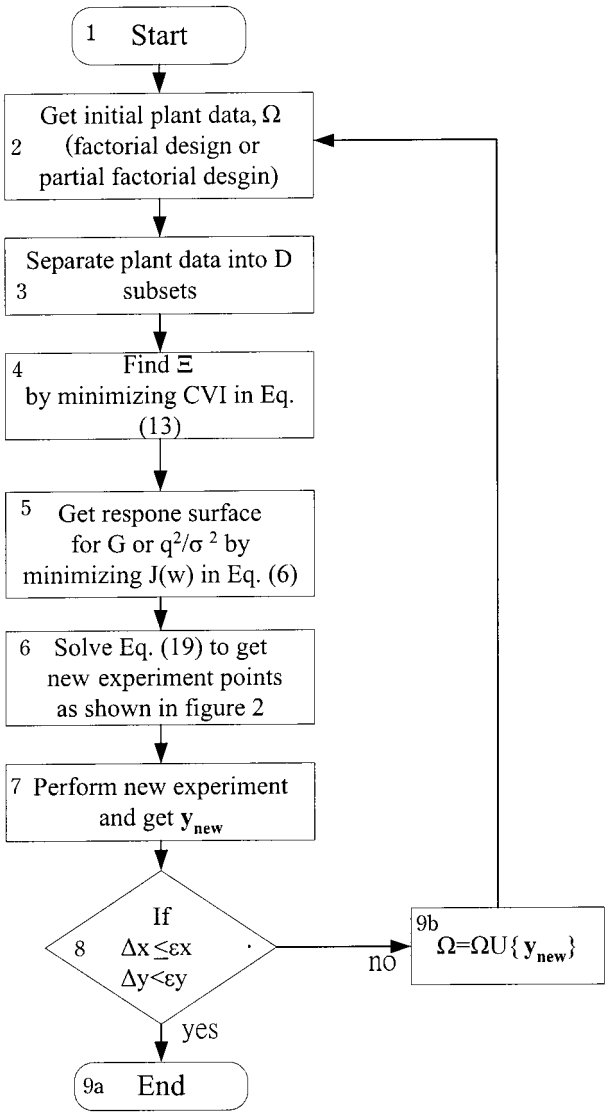

Figure 2. Flowchart of the ANN response surface experimental design approach.

the robust design in eq 19 is nontrivial because the response surface is usually nonconvex with multiple optima. The issues under such conditions have been thoroughly discussed in the authors' previous work.[1] We suggested an approach based on random search and information theory. In that approach, the response surface is investigated to find all probable optima. These optima are suggested as new experiments to be performed in the next run.

Figure 2 depicts the ANN response surface robust design approach proposed by this work. It can be divided into seven steps (steps 2−8 in Figure 2). Step 4 is the model determination stage, which implements the CVI approach discussed in section 2−2. Step 5 is the construction of the response surface stage, which determines the empirical model based on the training data. Step 6 is the optimization stage, which determines the possible optima based on the empirical model established in step 5. For the details of step 6, the reader is referred to Chen et al.[1]

3−2. Optimal Design Using the CVI-Based Model. As for the traditional response surface product design, the product quality q should be maximized. In most cases, a more general objective function that takes product cost, energy consumption, environment impact, and operational safety into consideration, $G(q,\mathbf{x})$ can be defined. The optimal design of a product or process can be formulated as

$$\max_{\mathbf{x}} \hat{G}(\hat{q},\mathbf{x})$$

$$\text{s.t.} \quad \hat{q} = \text{ANN}_q(\Xi,\mathbf{w},\mathbf{x}) \tag{20}$$

The methodology for solving eq 20 is the same as for the robust design problem in the previous section. If only the product quality objective is considered, then $\hat{G}$ can be replaced by $\hat{q}$. The efficiency of the optimization approach is highly improved, and the number of experiments is hence dramatically reduced, as shown in the next section and example 5−3.

4. The Pultrusion Process

Pultrusion is one of the most popular processes for manufacturing high-performance composite products. The flowchart in Figure 3 depicts a general pultrusion process. Fibers are stored and arranged in a Roving shelf. A puller is implemented to control the tension of the fibers to pass through a fiber guide and an impregnation tank; then the resin impregnated fibers go through the die. The die is heated, and the temperature is controlled. The products go through the puller, which also controls the pulling speed. A cutter is used to cut the products into a desired length. The mathematical models for temperature and composition profiles of this process have been well studied and are presented in section 4−1; a similar experimental setup is described in section 4−2.

4−1. Mathematical Model. Han,[33] Batch,[34] and Ma,[24] as well as their co-workers, have established the complete mathematical models for a pultrusion system to describe the monomer conversion and temperature distribution. In this section, the mathematical model derived by Han et al.[33] is briefly described. Note that a similar model[22] is also derived by the authors. The numerical specifications of simulated process are listed in Table 1. Figure 4 shows the Cartesian-coordinate model for a heating metal die.

(i) Continuity equation

$$V\frac{\partial C_A}{\partial z} = R_A \tag{21}$$

where V is the pulling speed, C_A denotes the concentration of resin, and R_A is the reaction rate of curing. Define the degree of curing as

$$\alpha = (C_{A0} - C_A)/C_{A0}$$

Then we have

$$\frac{d\alpha}{dt} = V\frac{\partial \alpha}{\partial z} = -\frac{V}{C_{A0}}\frac{\partial C_A}{\partial z} \tag{22}$$

(ii) Energy balance equation

$$\rho c_p V\frac{\partial T}{\partial z} = \frac{\partial^2 kT}{\partial x^2} + R_A\Delta H_R \tag{23}$$

where ΔH_R is the reaction heat; ρ is the bulk density; k is the bulk thermal conductivity; and C_{A0} is the initial

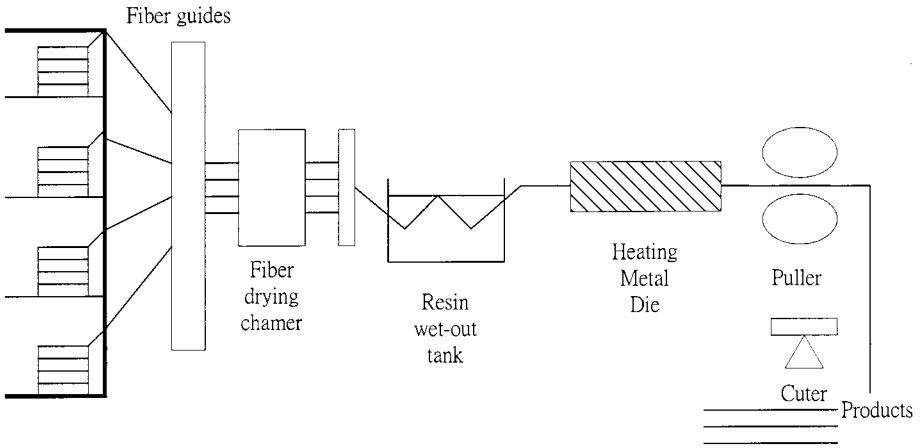

Figure 3. Schematic plot of the experimental setup.

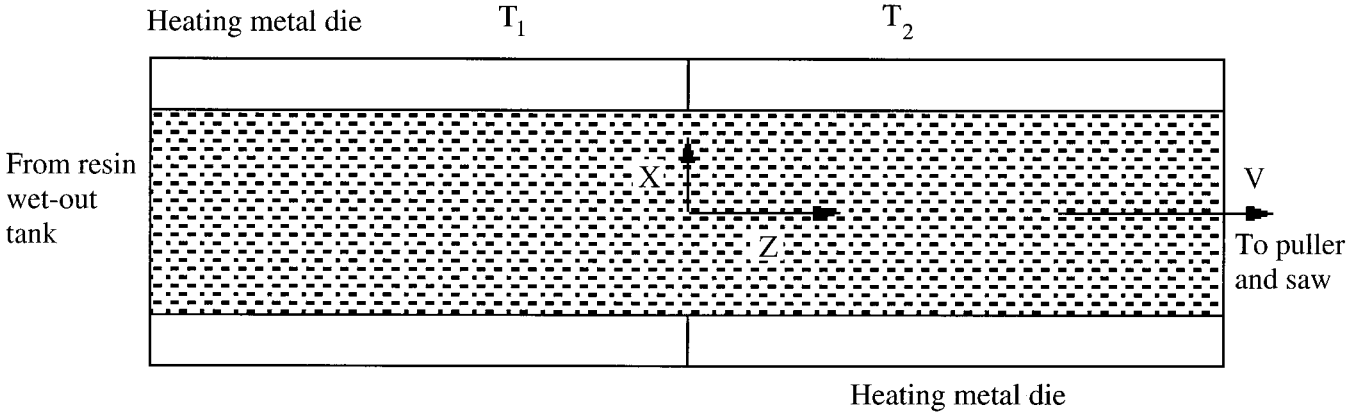

Figure 4. Coordinate and boundary conditions of the mathematical model for the heating metal die.

Table 1. Constants Used in Simulated Pultrusion Process

W, die width (cm)	1.27
L, die length (cm)	152.4
feed temperature (K)	333
fiber weight percentage	83
uncured polyester density (g/cm^3)	1.1
uncured polyester specific heat (cal g^{-1} K^{-1})	0.45
uncured polyester thermal conductivity (cal cm^{-1} s^{-1} K^{-1})	4.05×10^{-4}
cured polyester density (g/cm^3)	1.2
cured polyester specific heat (cal g^{-1} K^{-1})	0.45
cured polyester thermal conductivity (cal cm^{-1} s^{-1} K^{-1})	$1.54 \times 10^{-4} + 9.46 \times 10^{-7}T$
k_{10} (1/min)	3.412×10^{14}
E_1 (kcal/mol)	25.57
k_{20}(1/min)	5.167×10^{10}
E_2(kcal/mol)	17.93
m, order of reaction rate in eq 23	0.58
n, order of reaction rate in eq 23	1.42
ΔH_R, reaction heat (J/g)	200

concentration of reactive resin at the entrance, i.e., at $z = 0$.

(iii) Reaction rate

$$\frac{d\alpha}{dt} = (k_1 + k_2\alpha^m)(1 - \alpha)^n \qquad (24)$$

where

$$k_1 = k_{10}\exp(-E_1/RT) \quad \text{and} \quad k_2 = k_{20}\exp(-E_2/RT)$$

Table 2. Physical Properties of Glass Fiber and Resin Used in Experiments

(1) E-glass roving produced by PPG Industries, Inc., Pittsburgh, PA
density = 2.54 cm^3/g
fiber diamenter = 13 μm
tensile strength = $2-3 \times 10^5$
elastic modulus = 1.05×10^7

(2) PF-650 resole-type phenolic resin produced by Chunag-Tsung Co., Taiwan
viscosity = 150−250 mPa s
density (25 °C) = 1.1254 cm^3/g

The boundary conditions for eqs 20 and 22 can be listed as

$$x = L, \quad x = -L, \quad T = T(z)$$
$$z = 0, \quad T = T_i$$

4−2. Experimental Setup. In the laboratory, we used a bench-scale pultrusion system, Pulstar 810, Pultrusion Technology Incorp., as shown in Figure 3. The machine consists of two heating zones and two sets of pultrusion die with dimensions of 820 × 12.7 × 3.19 mm and 820 × 12.7 × 2.08 mm (length × width × thickness). The physical properties of the glass fiber and resin are listed in Table 2. In this system, two heating zones (T_1 and T_2), the temperatures of the mode, and the velocity (V) of the puller are controlled online. Therefore, three inputs, or designed variables, in eq 1 are available: $\mathbf{x} = (T_1, T_2, V)^T$. Three physical properties of the product are also measured. Fracture strength (F)

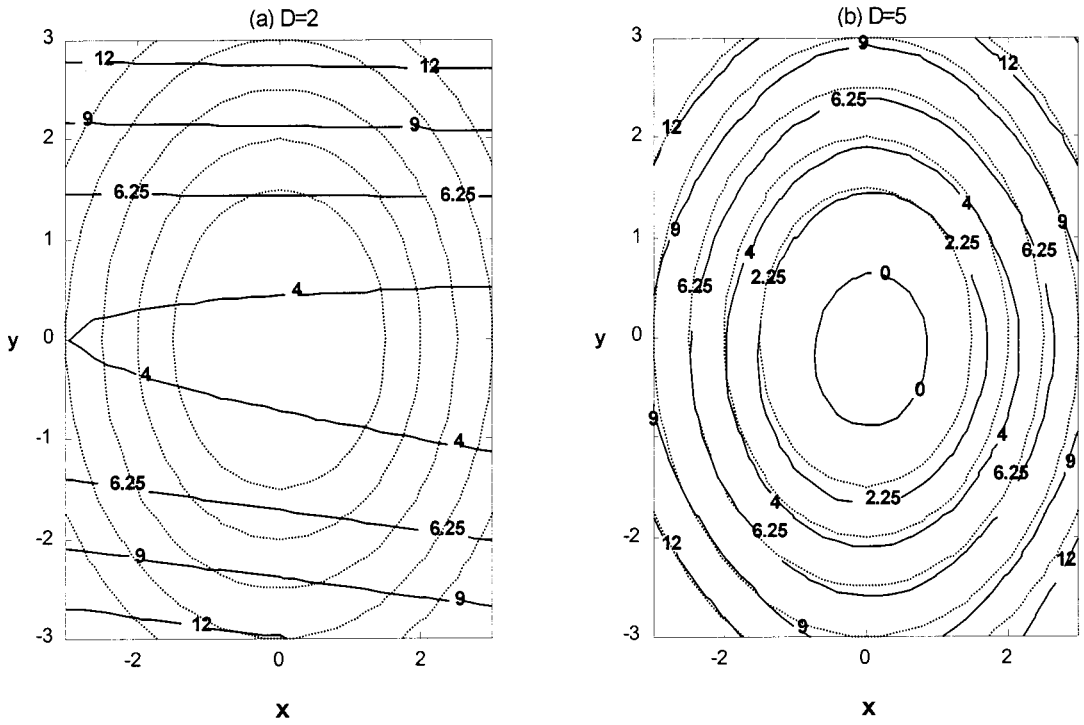

Figure 5. Comparison of the model behavior (solid lines) and the real quadratic function (dashed lines) based on CVI with 49 data points.

is measured using a universal testing machine (Instron Co.) following the specification of ASTM D-790. The sample dimensions are 12.6 × 1.21 × 0.29 cm. The span is 5 cm, and the crosshead speed is 2 mm/min. Tensile strength (T) is measured using an Instron-4468 instrument following the specification of ASTM D-3039. The sample dimensions are 25.6 × 1.21 × 0.29 cm, and the crosshead speed is 2.5 mm/min. The notched Irod impact strength (I) is measured using a TMI-43-1 machine (Testing Machine Inc.) according to ASTM D-256. The sample dimensions are 6.3 × 1.21 × 0.29 cm. The weight of the pendulum is 10 lb. In summary, the output variables in eq 1 are $\mathbf{y} = (T,F,I)^T$.

5. Results

In this section, we present the simulations and experiments of the pultrusion process described in the previous sections, as well as a simple curve fitting of a quadratic function to justify the theoretical work of the proposed cross-validation information index (CVI).

5−1. Curve Fitting of a Quadratic Function. The Effect of Data Dividing and Noise Size on CVI. According to proposition 2−4, increasing the number of divisions of the training set is desirable. However, the optimization problem in eq 13 needs to derive D sets of ANN submodels in every iteration. On the other hand, if the number of the experiments (size of the training set) increases, the solution of CVI would approach the exact solution. As mentioned before, the plant experiments are expensive. This example is to demonstrate the effects of divisions of the training set and size of noise in the training data to find appropriate model parameters using eq 13. Consider the quadratic function

$$q(x_1,x_2) = x_1^2 + x_2^2 + \epsilon$$

where $\epsilon \in N(0,2)$. Assume that 49 uniformly spaced data are sampled, and optimal CVIs in eq 13 are solved in the cases of $D = 2-6$. Notably, the subdivision of the above data set for cross-validation is performed randomly. Another separate testing set with a size of 961 is generated. Table 3 shows that $D = 5$ gives about the same results as $D = 6$ in terms of smoothness factor value, number of hidden nodes, and sum of absolute errors. The number of ANN submodels to be trained dramatically increases if D increases. Figure 5 compares the model predictions by optimal ANN structures based on $D = 2$ and 5 to the real curves. It is clear that, for the lower division of the training set, i.e., $D = 2$, the model is completely wrong.

Table 3. Effect of Subset Number on the Curving Fitting of a Quadratic Function

D	λ	hidden nodes	sum of absolute errors	number of neural network training iterations
2	0.0051	2	2500.3	84
3	0.0651	3	600.8	120
4	0.0575	3	498.7	176
5	0.0308	3	485.6	215
6	0.0341	3	473.8	312

Table 4. Effect of Noise Magnitude on the Curving Fitting of a Quadratic Function

noise distribution	λ	hidden nodes	sum of absolute errors
$N(0,2)$	0.0651	3	485.6
$N(0,5)$	0.07	3	1920.3
$N(0,8)$	0.09	3	4150.4

Table 4 shows that, in the case of higher noise magnitudes, this approach proposes higher smoothness factors. For the three cases, $\epsilon \in N(0,2)$, $\epsilon \in N(0,5)$, and $\epsilon \in N(0,8)$ in Table 4, the minimization of CVI proposed $\lambda = 0.0651$, 0.07, and 0.09, respectively, for a 49-data-point training set. The number of hidden nodes remains the same for all three cases. However, if the magnitude

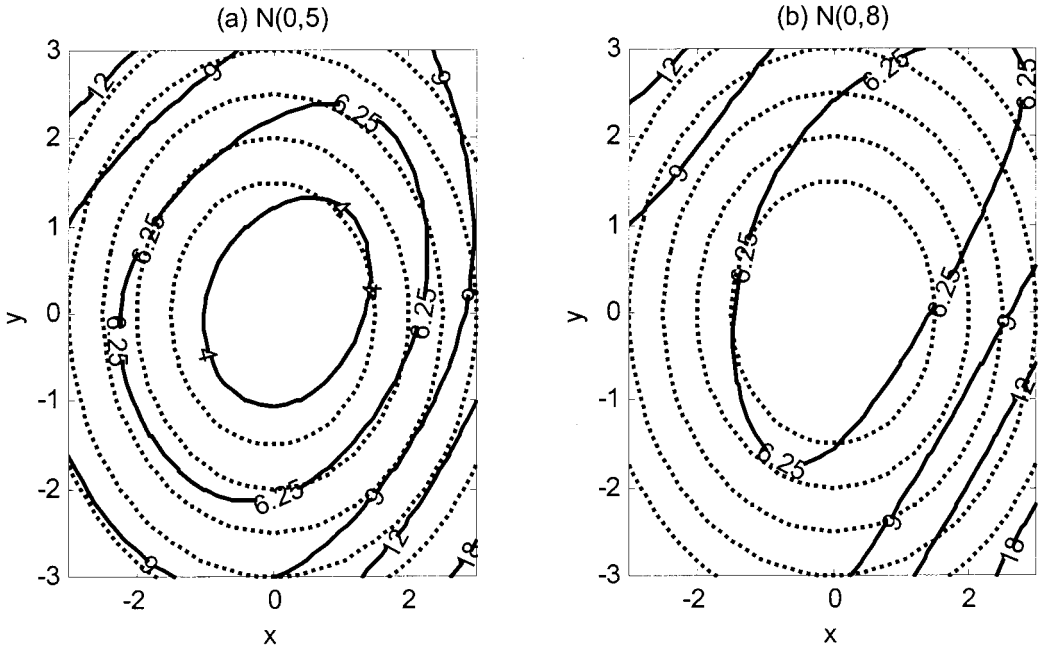

Figure 6. Comparison of the model behavior (solid lines) and the real quadratic function (dashed lines) based on data with different noise magnitude.

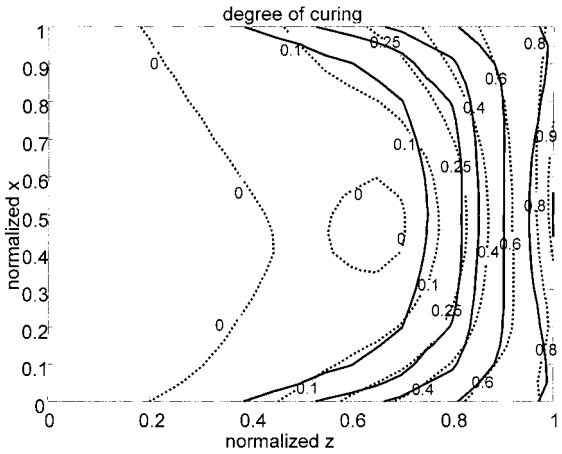

Figure 7. Curing profiles in the simulated die and CVI-based ANN model (solid lines, simulated data; dashed lines, ANN model).

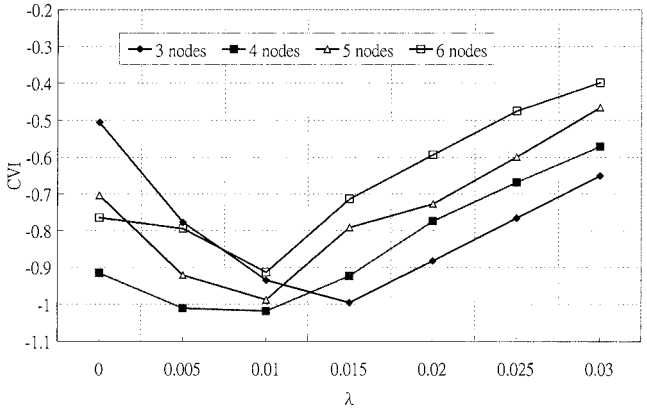

Figure 8. Determination of H and λ by CVI.

of noise becomes too high, i.e., $\epsilon \in N(0,8)$, the ANN model fails as shown in Figure 6. Increasing the number of experiments can solve this problem.

5−2. Simulation Study of the Pultrusion Process. The Study of Model Capability of CVI-Based ANN. Consider the mathematical model given in section 4−1 and assume that the model temperatures and pulling velocity are set at 353 K, 353 K, and 20 cm/s, respectively; then the partial differential equations 21 and 23 can be solved. In this work, the finite-difference technique suggested by both Han et al.[33] and Ma et al.[24] is implemented to solve the coupled PDEs. Figure 7 gives the extent of curing in the die. The curing profile $\alpha(x,z)$ in Figure 7 is modeled by using the CVI-based ANN developed in section 2, using 121 simulation points that are uniformly distributed and corrupted by normally distributed $N(0,0.02)$ noises. Figure 8 shows CVI values for different cases of ANN models for $\alpha(x,z)$ with varying numbers of hidden layer nodes and smoothness factors. The optimal settings of H and λ in eq 6 are found to be 4 and 0.01, respectively, as shown in Figure 8. A set of testing data with a size of 1331 points is used to test the accuracy of the acquired model. Figure 9 shows the total absolute errors between the prediction and testing data using ANN models with different numbers of hidden nodes. Figure 9 also compares the performances of the ANN models with and without smoothness factors. It is obvious that the ANN models with large nnumbers of hidden layer nodes are overfitted because of the effect of noises and limited data. The results indicate that the ANN model structure and smoothness factor are determined by CVI appropriately. Table 5 compares the optimal model structures determined by all existing techniques and our proposed index. CVI is superior to all other approaches in the cases of $Q = 121$ and 289. In the case of $Q = 441$, CVI and all other approaches give the same model paraemters because the number of experiments is sufficient for a smooth regression. It should be noted that the optimal smoothing factor λ based on CVI decreases to zero as the size of the training set increases. This is consistent with the theoretical analyses in section 2.

5−3. Experimental Study. Consider the experiments of the pultrusion process described in section 4−2; the maximum tensile strength, fracture strength,

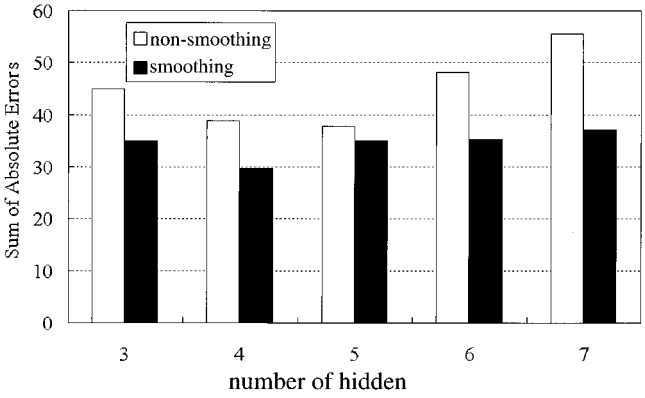

Figure 9. Sum of absolute errors between the model predictions and testing data using different model structures with/without an optimal smoothness factor.

Table 5. Optimal ANN Structure Determined by Various Statistical Indices in the Simulated Pultrusion Process

number of data points	indices	number of hidden nodes	λ	sum of absolute errors
121	CVI	4	0.01	27.3451
	cross-validation	6	0.01	32.5869
	AIC	6	0.005	31.9238
	conditional AIC	4	0.015	33.9053
289	CVI	5	0.0075	22.0165
	cross-validation	6	0.0025	24.1933
	AIC	6	0.001	24.8509
	conditional AIC	6	0.012	23.6843
441	CVI	6	0.00	18.7088
	cross-validation	6	0.00	18.7088
	AIC	6	0.00	18.7088
	conditional AIC	6	0.00	18.7088

and impact strength are sought. Therefore, we define the product mechanical property objective as

$$q = \frac{T - T_{\min}}{T_{\max} - T_{\min}} + \frac{F - F_{\min}}{F_{\max} - F_{\min}} + \frac{I - I_{\min}}{I_{\max} - I_{\min}} \quad (25)$$

where subscripts max and min denote, respectively, the maximum and minimum obtainable values of product quality in tensile strength, fracture strength, and impact strength. In addition, the product capacity, i.e., speed of pulling, is also concerned. Consequently, the overall production objective is formulated as

$$\max_{T_1,T_2,V} G = V + 20\hat{q}$$

$$\text{s.t.} \quad \hat{q} = \text{ANN}_q(T_1,T_2,V) \quad (26)$$

where 20 is used to scale up the magnitude of the above mechanical property objective, q. In most industry cases, an orthogonal array table or partial factorial design is preferable. Here, for a thorough study, 27 independent experiments were performed based on a full factorial design, as shown in Table 6. For each experiment, five samples are cut from the products and the physical properties (T, F, I) were measured. The box plots for these 27 experiments are shown in Figure 10a−10c. The noises of these product quality measurements are very large.

Table 7 shows the results of conducting an ANOVA (analysis of variance) for the product quality. All factors (T_1, T_2, and V) and their interaction terms have strong effects[27] (low P values) on product quality. The system

Table 6. Results of Initial Full Factorial Design for Experimental Pultrusion Process

	T_1 (°C)	T_2 (°C)	V (cm/min)	$E(q)$	$\log_{10}(q^2/\sigma^2)$	σ^2	G
1	200	200	10	1.060	1.352	0.050	31.2
2	200	200	20	0.220	−0.530	0.164	24.4
3	200	200	30	0.298	0.907	0.011	35.8
4	200	220	10	1.555	1.605	0.060	41.2
5	200	220	20	0.926	1.571	0.023	38.6
6	200	220	30	0.865	1.833	0.011	47.2
7	200	240	10	1.961	1.052	0.341	44.4
8	200	240	20	1.178	1.426	0.052	43.4
9	200	240	30	0.963	1.853	0.013	49.4
10	220	200	10	1.368	1.841	0.027	37.4
11	220	200	20	1.483	1.643	0.050	49.6
12	220	200	30	0.729	1.612	0.013	44.6
13	220	220	10	1.844	1.841	0.049	46.8
14	220	220	20	1.074	2.108	0.009	41.4
15	220	220	30	0.602	0.957	0.040	42
16	220	240	10	1.705	1.314	0.141	44.2
17	220	240	20	1.270	1.408	0.063	45.2
18	220	240	30	0.973	2.073	0.008	49.4
19	240	200	10	1.429	1.328	0.096	36
20	240	200	20	0.940	1.107	0.069	38.8
21	240	200	30	1.025	1.590	0.027	50.4
22	240	220	10	2.038	1.780	0.069	48.6
23	240	220	20	0.796	1.459	0.022	36
24	240	220	30	0.711	0.911	0.062	44.2
25	240	240	10	1.975	1.792	0.063	51.4
26	240	240	20	0.528	0.998	0.028	30.6
27	240	240	30	0.774	0.552	0.168	44.4

is highly nonlinear. A traditional statistical cubic function model

$$\hat{q} = -21.32 + 0.16025T_1 + 0.03483T_2 - 1.6311V + 0.00022T_1T_2 + 0.007T_2V + 0.0072T_1V - 0.000452{T_1}^2 - 0.000149{T_2}^2 + 0.00257V^2 - 3.406 \times 10^{-5}T_1T_2V \quad (27)$$

is obtained by regressing the 27 experimental data points. This model is, in turn, implemented to find optimal operating conditions, as shown in Table 8. The best product quality, G, in Table 8 is 48.67, which is smaller than the values obtained in at least five runs in the initial factorial design shown in Table 6. The result shows that the above model fails to find better operating conditions.

A normal probability plot for the signal-to-noise ratio (S/N), shown in Figure 11, indicates that the S/N ratio is clearly related to the operating conditions. In Figure 11, strong effects of all factors and their interactions on the S/N ratio are indicated by their significant deviations from the dashed line.[27] Therefore, in terms of the S/N ratio, the system is also highly nonlinear.

Model Capability. Using the results of the 27 experiments, the means of the three physical properties are calculated, and the response surface of the product quality objective $\hat{q}$ is constructed by solving the double optimization problem in eqs 6 and 13. The optimal number of the nodes, H, in the hidden layer is 4, and the smoothness factor ($\lambda = 0.03$) is also found. Figure 12a shows the contour plot of the CVI-based ANN model for the product quality by only selecting the effects of variables T_1 and T_2. This contour plot shows that at least two local optima exist in the operating ranges of T_1 and T_2. Hence, this system has multiple optima and is nonconvex. However, if one implements an ANN model with the same structure as the previous model without a smoothness factor, then the contour plot given

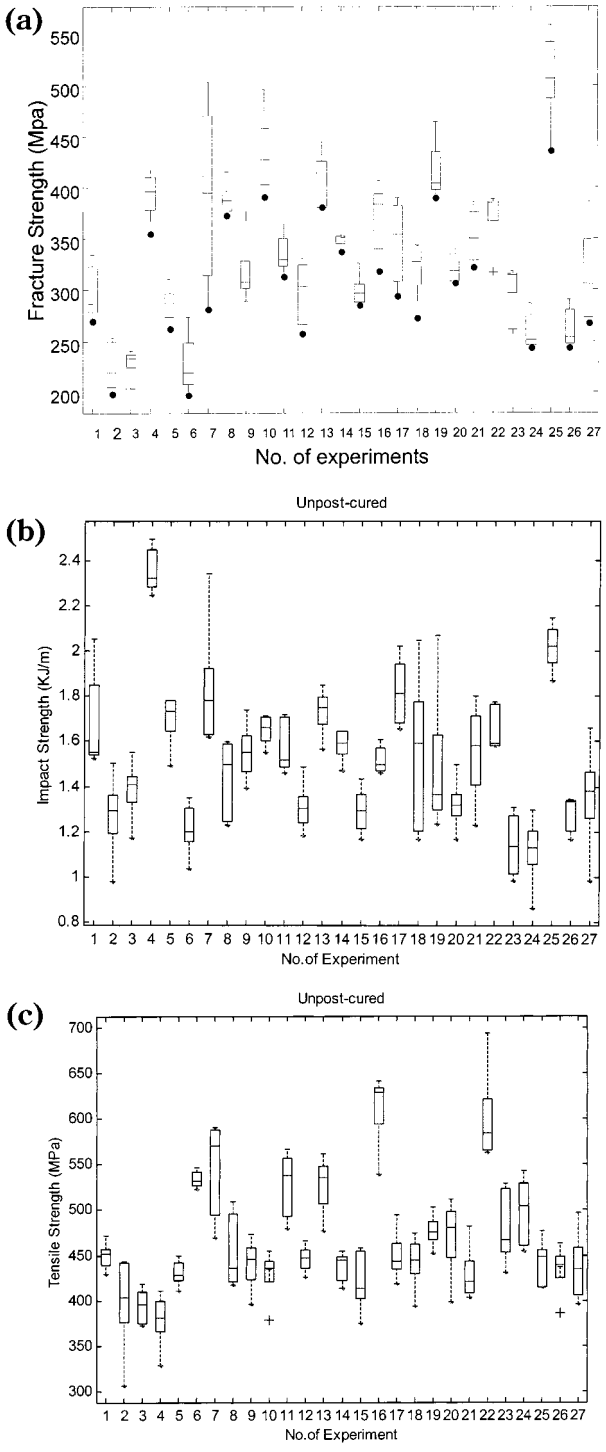

Figure 10. Product span of 27 independent experimental runs: (a) fracture strength, (b) impact strength, (c) tensile strength.

in Figure 12b is obtained. The contour plot becomes very irregular and impossible to implement.

For the purpose of comparison, AIC and CV were also implemented to determine an ANN model with a smoothing term using the 27 training data points in Table 6. The number of hidden layer nodes (H) is 5 for AIC and 5 for CV, while the smoothing factor (λ) is 0.01 for AIC and for CV (compared with $H = 4$ and $\lambda = 0.03$ for CVI). The first four experimental data points in run I of Table 9 are used as the testing data. The testing errors are shown in Figure 13. The results show that the CVI-based model is much more reliable than the AIC- and CV-based models.

Table 7. ANOVA Analysis for Quality Variable to the Design Variables

source of variation	SS	df	MS	Fo	*P* value
T_1 (A)	1.1484	2	0.5742	9.2829	0.0002
T_2 (B)	2.2211	2	1.1105	17.9543	0
V (C)	20.1164	2	10.0582	162.6128	0
AB	3.6081	4	0.902	14.5831	0
AC	2.2221	4	0.5555	8.9814	0
BC	1.457	4	0.3642	5.8887	0.0004
ABC	1.7101	8	0.2138	3.456	0.0036
error	6.6802	108	0.0619		
total	39.1633	134			

Table 8. Results of Traditional Experimental Design with a Second-Order Polynomial Model

	T_1 (°C)	T_2 (°C)	*V* (cm/min)	*q*	*q* (predicted)	error (%)	*G*
I	203.12	240	30	0.933534	1.048	12.26158	48.67068

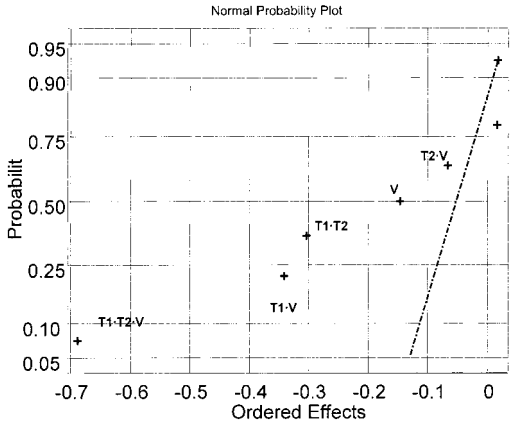

Figure 11. Normal probability plot of operating condition effects on *S*/*N* ratio.

Table 9. Results of ANN Experimental Design with Smooth Training Neural Model

run		T_1 (°C)	T_2 (°C)	*V* (cm/min)	*q*	*q* (predicted)	error (%)	*G*
I	1	**209.1**	**231.46**	**12.05**	**1.99**	**1.88**	**−5.68**	**51.85**
	2	**239.27**	**236.55**	**11.00**	**1.89**	**1.84**	**−2.7**	**48.8**
	3	228.5	226.79	29.53	0.9	0.91	1.33	47.53
	4	217.24	208.96	20.84	1.24	1.2	−3.38	45.64
II	1	**210.24**	**239.04**	**14.25**	**1.900**	**1.82**	**−4.195**	**52.24**
	2	**231.09**	**239.5**	**13.3**	**1.848**	**1.78**	**−3.684**	**50.26**
	3	231.18	228.57	12.73	1.788	1.66	−7.146	48.49
	4	207.43	229.49	26.62	1.009	1.14	13.02	46.79
	5	230.29	202.62	21.49	1.068	1.14	6.732	42.85
	6	210.88	210.57	26.99	0.974	1.11	13.97	46.47
III	1	**208.27**	**239.25**	**14.55**	**2.044**	**1.87**	**−8.525**	**55.44**
	2	**229.07**	**237.26**	**12.23**	**1.863**	**1.86**	**−0.139**	**49.48**
	3	211.74	221.19	13.72	1.811	1.87	3.261	49.97
	4	213.8	216.1	28.89	0.929	1.07	15.208	47.46
	5	225.54	207.16	28.22	0.962	1.07	11.237	47.46

Optima Design. Using the acquired CVI-based ANN quality model, the optimization algorithm described in Figure 2 is performed. The optimization problem in eq 26 is solved. Our information theory and random search algorithm[1] suggests several new experiments in each run. The results of three runs of the optimization procedures for product design are documented in Table 9. In this particular system, as indicated by Figure 12a, two local optima can be found for the objective function denoted by eq 26. The bold rows of Table 9 record the evolutionary history of two optimal conditions through three iterations of the proposed optimization procedure. The suggested experimental conditions converge to the optima. It should be noted that our optimization ap-

Table 10. Results of ANN Experimental Design without Smooth Training Neural Model

run		T_1 (°C)	T_2 (°C)	V (cm/min)	q	q (predicted)	error (%)	$V + 20q$
I	1	230.35	208.87	15.82	1.5475	1.76	13.73179	46.77001
	2	229.57	222.06	19.56	1.537289	0.83	−46.0088	50.30577
	3	208.01	230.77	16.24	1.308228	1.71	30.71115	42.40456
	4	201.31	219.51	16.55	1.847937	1.76	−4.75866	53.50874
	5	211.46	211.42	16.72	0.825945	1.73	109.4571	33.2389
II	1	228.14	229.44	17.8	1.34273	1.16	−13.6088	44.6546
	2	205.37	232.41	21.44	1.165695	4.47	283.4621	44.75391
	3	212.14	209.66	20.3	1.156539	2.93	153.342	43.43078
III	1	234.93	234.07	14.27	1.416319	2.29	61.68677	42.59638
	2	210.45	232.18	21.44	1.33883	1.62	21.00119	48.21657
	3	225.17	212.24	19.9	1.515493	0.99	−34.6747	50.20986

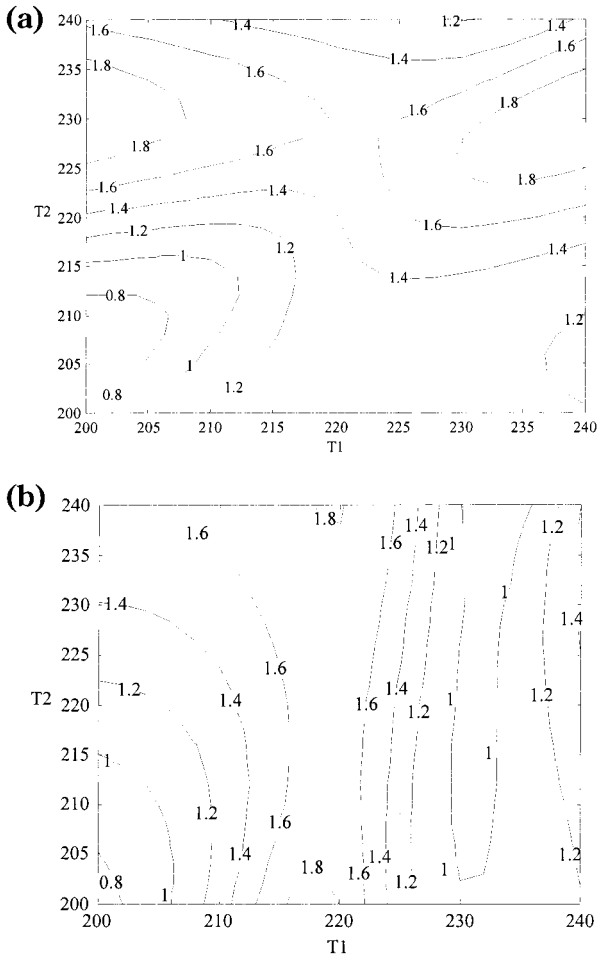

Figure 12. Artificial neural network model response surfaces determined by experiments: (a) ANN model with smooth training, (b) ANN model without smooth training.

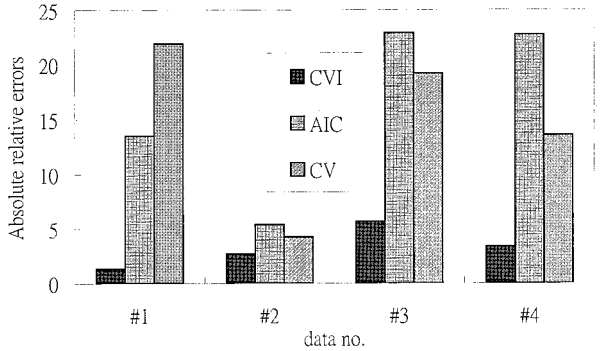

Figure 13. Comparison of the testing errors of three ANN models whose structures are determined by CVI, AIC, and CV.

proach[1] usually suggests more than one new condition to the user. This is because of our random search clustering approach. For details, the reader is referred to Chen et al.[1] That is why four to six experiments including the two optima (the bold rows in Table 9) are conducted in each run.

Table 10 shows the same case as Table 9. The only difference is that we implement nonsmoothing ANN model, as shown in Figure 12b. Moreover, the experimental points suggested in runs 2 and 3 are still far away from the final optimal operating conditions suggested by the smoothing response surface approach. By comparing Figure 12 and Tables 9 and 10, it is very clear that the implementation of a CVI-based model with the consideration of smoothness factor is highly promising for finding the optimal operating conditions in a highly noisy system with a limited number of experimental data.

Robust Design. Assume that it is also desirable to find a robust operating condition (high S/N ratio in eq 19) for the same pultrusion process. Using the experimental results shown in Table 6, a CVI-based ANN model with a smoothing term can be derived. Subsequently, the robust design problem in eq 19 can be solved as the previous case, i.e., by replacing the objective function G in eq 26 by $\hat{q}^2/\sigma^2$. Table 11 shows that robust operating conditions can be found. Comparing the optimal result with the original 27 experiments in Table 6, this robust design approach significantly improves the S/N ratio from the full factorial design results. It should be noted that the original Taguchi method can only find the best operating condition from the above 27 data points. According to Table 11, two optima, items 1 and 2 in the third batch, are listed. One is at $T_1 = 229$, $T_2 = 232$, and $V = 17$; the other is at $T_1 = 214$, $T_2 = 238$, and $V = 28$. In the first case, the product quality is good, and the variance is comparatively large than it is in the second case. In the second case, the product quality is not as good as in the first case, but the product variance is very low. This is because, at high temperature, the curing of the product in the die is more complete. However, because of the effect of the curing, the high-temperature products become not as uniform in flow as the low-temperature products. There exists a compromise between the two cases. Finally, it should be noted that the proposed approach can handle multiple-optima problems that cannot be treated by the traditional response surface method (RSM).

6. Conclusion

A novel approach to determining the model structure and regression parameters of an ANN response surface model is developed. The novel information index (cross-validation information index) is a combination of the

Table 11. Results of Robust Experimental Design

run		T_1 (°C)	T_2 (°C)	V (cm/min)	q	var	$\log_{10}(q^2/\sigma^2)$ (predicted)	error (%)
I	1	228.53	232.69	17.49	2.1223	0.0072 (2.7012)	2.79639	−3.3988
	2	211.61	236.71	28.95	1.2378	0.0057 (2.699)	2.4294	11.0942
	3	227.79	209.57	12.02	1.5476	0.0325 (1.7402)	1.8674	−6.8137
	4	201.41	218.86	13.30	1.3412	0.0088 (2.5811)	2.3105	11.7116
II	1	230.18	231.41	15.81	1.7611	0.0094 (2.6947)	2.5184	6.9998
	2	213.15	235.95	28	1.4929	0.0046 (2.731)	2.6853	1.7011
	3	227.02	210.02	12.29	1.5492	0.0553 (1.8182)	1.6375	11.0377
	4	201.28	213.59	12	1.2501	0.0055 (2.3912)	2.4535	−2.5399
III	1	229.49	232.19	17.13	1.7802	0.0081 (2.5603)	2.5925	−1.2419
	2	213.82	237.58	27.97	1.4395	0.0039 (2.6669)	2.7254	−2.1449
	3	228.28	210.41	13.38	1.5490	0.0445 (1.6778)	1.7317	−3.1131
	4	201.15	214.3	12.04	1.2095	0.0087 (2.4421)	2.2257	9.7239

Akaike information index and cross-validation. The theoretical analysis shows that the new index converges rapidly when a limited number of noisy experimental data are available. This approach is, in turn, implemented to find the optimal/robust design solution of a pultrusion process. Both the simulation and the experimental works show that the proposed approach is very promising.

Acknowledgment

The authors acknowledge the financial support provided by the National Science Council through Grant NSC-88-CPC-E007-015.

Appendix

Proof of Proposition 2−2. The following proof is a natural extension from AIC except that the likelihood function is replaced by the conditional likelihood function in eq 12. For details, the reader is referred to Tong.[17] Given the data set x of Q observations, the prediction is realized by $g(y|x)$, and the true distribution of y is noted as $f(y)$. The Kullback−Leibler information entropy can be solved as

$$I(f,g) = -\int f(y) \ln\frac{g(y|x)}{f(y)}\,dy = -\int \frac{f(y)}{g(y|x)} \ln\left(\frac{g(y|x)}{f(y)}\right) g(y|x)\,dy \quad \text{(A-1)}$$

$$= \int f(y) \ln g(y|x)\,dy - \int f(y) \ln f(y)\,dy$$

$$= E_y \ln g(y|x) - \text{constants}$$

The goodness of the prediction of a future observation is defined as $E_xE_y \ln g(y|x)$. Suppose that x and y are independent and that $g(y|x) = g(y|\Xi)$ is a distribution specified by a set of parameters Ξ. The conditional logarithm of the likelihood of the data-dependent model $g(y|x)$ is defined by

$$\ln g(y|\Xi) + C$$

where $C = \dim[\Xi]$ is a constant correction term such that

$$E_x[\ln g(y|\Xi) + C] = E_xE_y \ln g(y|x) \quad \text{(A-2)}$$

(1) Suppose that the optimal setting of Ξ is Ξ^* and that $\hat{\Xi}$ is the conditional maximum likelihood estimation when $\dim[\Xi]$ is previously determined; then the conditional likelihood ratio statistic has the asymptotic x_p^2 distribution.

$$\lim_{Q\to\infty} 2 \ln g(x|\Xi^*) - 2 \ln g(x|\hat{\Xi}) \approx \chi_p^2, \quad p = \dim[\Xi,w] - 1 \quad \text{(A-3)}$$

(2) By expanding $2 \ln g(y|\hat{\Xi})$ in the neighborhood of $2 \ln g(y|\Xi^*)$ and ignoring the higher-order terms, we have

$$\sqrt{Q}(\hat{\Xi} - \Xi^*)^T \approx N(0, I_{\Xi^*}^{-1}), \quad \text{as } Q \to \infty$$

$$I_{\Xi^*} = -\frac{1}{Q}E_y\left[\frac{\partial^2 \Xi^*}{\partial \Xi^2}\right] \quad \text{(A-4)}$$

This completes the proof of proposition 2−3.

Note that

$$2E_y[\ln g(y|\Xi^*) - 2 \ln g(y|\hat{\Xi})] = \sqrt{Q}(\hat{\Xi} - \Xi^*)^T I_{\Xi^*} \sqrt{Q}(\hat{\Xi} - \Xi^*)] \approx \chi_p^2$$

Combining eqs A-3 and A-5, we have

$$2E_x[2 \ln g(y|\hat{\Xi}) - \ln g(y|\Xi^*)] = p$$

and

$$2E_xE_y[\ln g(y|\Xi^*) - 2 \ln g(y|\hat{\Xi})] = p$$

We can thus conclude that $C = -p$.

Now, because the conditional likelihood function can be defined as

$$L = \prod_{i=1}^{D} L_i = \left(\frac{1}{2\pi\sigma^2}\right)^{-N/2} \exp\left[-\frac{1}{2\pi\sigma^2} \sum_{i=}^{D} \sum_{x_i \in \Omega_i} (y(x_i) - \hat{\mathbf{y}}(x_i))^2 \right]$$

where N_i is the number of elements in Ω_i, the CVI can be written as

$$\mathrm{CVI} = -2 \ln L + 2p = Q \ln(2\pi) + Q \ln(\sigma^2) + Q + 2p$$

Rearranging the above equation and neglecting constants, we obtain

$$\mathrm{CVI} = \ln(\sigma^2) + \frac{2}{Q}p = \ln\left(\frac{1}{N}\sum_{i=}^{D}\sum_{x\in\Omega_i}(y(x) - \hat{y}(x))^2\right) + \frac{2}{Q}p \tag{A-7}$$

Proof of Proposition 2−3. Because $\Omega = \{(\mathbf{x}_i,\mathbf{y}_i)|i = 1, ..., Q\}$ is a set of Markov-dependent observations, the joint probability density function of $\{x_1,...,x_q\}$ is given by

$$p(x_1,...,x_q) = p(x_1)\, p(x_2|x_1) \cdots p(x_q|x_{q-1})$$

Suppose that θ^* is the optimal setting; then

$$d_\theta \mathrm{CVI}(\theta^*) = 0$$

where d_θ denotes the operator of the first derivatives with respect of θ. The usual Taylor expansion of CVI(θ) about θ^* would be

$$0 = d_\theta \mathrm{CVI}(\theta^*) = d_\theta \mathrm{CVI}(\hat{\theta}) + d_\theta{}^2 \mathrm{CVI}(\hat{\theta})(\theta^* - \hat{\theta}) + O(\theta^* - \hat{\theta})$$

If the last residual term in the above equation is neglected, then

$$\sqrt{Q}(\theta^* - \hat{\theta}) = -\left(\frac{1}{Q} d_\theta{}^2 \mathrm{CVI}(\hat{\theta})\right)^{-1}\left(\frac{1}{\sqrt{Q}} d_\theta \mathrm{CVI}(\hat{\theta})\right)$$

According to Martingale theory[35]

$$\lim_{Q\to\infty}\left(\frac{1}{Q} d_\theta{}^2 \mathrm{CVI}(\hat{\theta})\right) = \mathbf{V}$$

where $\mathbf{V}$ is a constant matrix in probability. This simply requires ergodicity, and the stationarity of $\Omega = \{(\mathbf{x}_i,\mathbf{y}_i)|i = 1, ..., Q\}$ is obtained. By the Martingale central limit theorem[17]

$$\lim_{Q\to\infty}\left(\frac{1}{\sqrt{Q}} d_\theta \mathrm{CVI}(\hat{\theta})\right) = \mathbf{W}\mathbf{V}^{-1}$$

where $\mathbf{W}$ and $\mathbf{M}$ are constant matrices.

$$\sqrt{Q}(\theta^* - \hat{\theta}) \approx N(0,\mathbf{M}), \mathbf{M} = \mathbf{V}^{-1}\mathbf{W}\mathbf{V}^{-1} \quad \text{as } Q \to \infty$$

Proof of Proposition 2−4.

$$\mathrm{CVI} = \ln(\sigma^2) + \frac{2}{Q}p = \ln\left(\frac{1}{Q}\sum_{i=}^{D}\sum_{x_i\in\Omega_i}(y(x_i) - \hat{y}(x_i))^2\right) + \frac{2}{Q}p$$

$$\hat{\theta}_D = \arg\min_{\Xi}\left[\ln\left(\frac{1}{Q}\sum_{i=}^{D}\sum_{x_i\in\Omega_i}(y(x_i) - \hat{y}(x_i))^2\right) + \frac{2}{Q}p\right]$$

where m is an integer such that $m = [Q/D]$ ([x] denotes the largest integer not greater than x).

Denote the residual, $(\hat{y} - \bar{y})$, as $\epsilon(x,\theta)$. Then, we can write a Taylor expansion as

$$\epsilon^2(x,\hat{\theta}_D) = \epsilon^2(x,\theta^*) + 2\epsilon(x,\theta^*) d_\theta \epsilon(\hat{\theta}_D - \theta^*) + O(|\hat{\theta}_D - \theta^*|)$$

$$\begin{aligned} 0 &= \frac{\partial}{\partial\theta}\left[\frac{1}{Q}\sum_{x\in\Omega-\Omega_i}\epsilon^2(x,\theta)\right]_{|\theta=\hat{\theta}_p} \\ &= d_\theta V(\hat{\theta}_D) - \frac{2}{Q}\sum_{x\in\Omega_i}\epsilon(x,\hat{\theta}_D) d_\theta \epsilon(x,\hat{\theta}_D) \\ &= d_\theta V(\theta^*) - \frac{2}{Q}\sum_{x\in\Omega_i}\epsilon(x,\theta^*) d_\theta\epsilon(x,\theta^*) + \left\{ d_\theta{}^2 V(\theta^*) - \frac{m}{Q}\frac{\partial}{\partial\theta}\left[\frac{2}{m}\sum_{x\in\Omega_i}\epsilon(x,\theta) d_\theta\epsilon(x,\theta)\right]\Big|_{\theta=\theta^*}\right\}(\hat{\theta}_D - \theta^*) + O(|\hat{\theta}_D - \theta^*|^2) \end{aligned} \tag{A-8}$$

where

$$V(\theta) = \sum_{x\in\Omega-\Omega_i}\frac{1}{Q}\|\hat{\mathbf{y}}^i(x,\theta) - \bar{\mathbf{y}}^i(x,\theta)\|^2 + \frac{2}{Q}p$$

Because (E denotes expectation)

$$\begin{aligned} \frac{1}{m}\sum_{x\in\Omega_i}\epsilon(x,\theta^*) d_\theta\epsilon(x,\theta^*) &= E\sum_{x\in\Omega_i}\epsilon(x,\theta^*) d_\theta\epsilon(x,\theta^*) + O\left(\frac{1}{\sqrt{m}}\right) \\ &= \left\lfloor \frac{1}{Q}\sum_{x\in\Omega}\epsilon(x,\theta^*) d_\theta\epsilon(x,\theta^*) + O\left(\frac{1}{\sqrt{Q}}\right)\right\rfloor + O\left(\frac{1}{\sqrt{m}}\right) \end{aligned} \tag{A-9}$$

it follows from eq A that

$$|\hat{\theta}_D - \theta^*| = O\left(\frac{1}{D\sqrt{m}}\right) = O\left(\frac{1}{\sqrt{D}\sqrt{Q}}\right)$$

Literature Cited

(1) Chen, J.; Wong, D. S. H.; Jang, S. S.; Yang, S. L. Product and Process Development Using Artificial Neural Network Model and Information Theory. *AIChE J.* **1998**, *44*, 876.

(2) Doherty, S. K.; Gomm, J. B.; Willems, D. Experiment Design Considerations for Nonlinear System Identification Using Neural Networks. *Comput. Chem. Eng.* **1997**, *21*, 327.

(3) Su, H. T.; McAvoy, T. J.; Werbos, P. Long-Term Predictions of Chemical Processes Using Recurrent Neural Networks: A Parallel Training Approach. *Ind. Eng. Chem. Res.* **1992**, *31*, 1338.

(4) Joseph, B.; Hanratty, F. W. Predictive Control of Quality in a Batch Manufacturing Process Using Artificial Neural Network Models. *Ind. Eng. Chem. Res.* **1993**, *32*, 1951.

(5) Sridhar, D. V.; Seagrave, R. C.; Bartlett, E. B. Process Modeling Using Stacked Neural Networks. *AIChE J.* **1996**, *42*, 2529.

(6) Lee, M.; Park, S. A New Scheme Combining Neural Feedward Control with Model-Predictive Control. *AIChE J*. **1992**, *38*, 193.

(7) Chen, Q.; Weigand, W. A. Dynamic Optimization of Nonlinear Processes by Combining Neural Net Model with UDMC. *AIChE J*. **1994**, *40*, 1488.

(8) Nahas, E. P.; Henson, M. A.; Seborg, D. E. Nonlinear Internal Model Control Strategy for Neural Network Models. *Comput. Chem. Eng.* **1992**, *16*, 1039.

(9) Shaw, A. M.; Doyle, F. J., III. Multivariable Nonlinear Control Applications for a High Purity Distillation Column Using a Recurrent Dynamic Neuron Model. *J. Process Control* **1997**, *7*, 255.

(10) Psichogios, D. C.; Ungar, L. H. A Hybrid Neural Network−First Principles Approach to Process Modeling. *AIChE J*. **1992**, *38*, 1499.

(11) Thompson, M. L.; Kramer, M. A. Modeling Chemical Processes Using Prior Knowledge and Neural Networks. *AIChE J*. **1994**, *40*, 1328.

(12) Karjala, T. W.; Himmelblau, D. M. Dynamic Rectification of Data via Recurrent Neural Nets and the Extended Kalman Filter. *AIChE J*. **1996**, *42*, 2225.

(13) Tsen, A. Y. D.; Jang, S. S.; Wong, D. S. H.; Joseph, B. Predictive Control of Quality in Batch Polymerization Using a Hybrid Artificial Neural Network Model. *AIChE J*. **1996**, *42*, 455.

(14) Lin, J.-S.; Jang, S.-S.; Shieh, S.-S.; Subramaniam, M. Generalized Multivariable Dynamic Artificial Neural Network Modeling for Chemical Processes. *Ind. Eng. Chem. Res.* **1999**, *38*, 12, 4700.

(15) Murata, N.; Yoshizawa, M.; Amari, S. I. Network Information Criterion−Determining the Number of Hidden Units for an Artificial Neural Network Model. *IEEE Trans. Neural Networks* **1994**, *5*, 865.

(16) Fogel, D. B. An Information Criterion for Optimal Neural Network Selection. *IEEE Trans. Neural Networks* **1991**, *2*, 490.

(17) Tong, H. *Nonlinear Time Series: A Dynamical System Approach*; Oxford University Press: New York, 1990.

(18) Stoica, P.; Ekyhoff, P.; Janssen, P.; Soderstrom, T. Model-Structure Selection by Cross-Validation. *Int. J. Control* **1986**, *43*, 1841.

(19) Galatsanos, N.; Katsaggelos, A. K. Cross-Validation and Other Criteria for Estimating the Regularizing Parameter. *IEEE Trans. Acoust., Speech, Signal Process.* **1991**, *4*, 3021.

(20) Moody, J. E. Note on Generalization, Regularization and Architecture Selection in Nonlinear Learning Systems. In *Neural Networks for Signal Processing, Proceedings of the 1991 IEEE Workshop*; IEEE: New York, 1991; Vol. 1.

(21) Poggio, T.; Girosi, F. Regularization Algorithms for Learning that Are Equivalent to Multilayer Networks. *Science* **1990**, *247*, 978.

(22) Amari, S. N.; Murata, K.; Müller, K. R.; Finke, M.; Yang, H. H. Asymptotic statistical Theory of Overtraining and Cross-Validation. *IEEE Trans. Neural Networks* **1997**, *8*, 985.

(23) Ma, C. C. M.; Chen, C. C. The Development of a Mathematical Model for the Pultrusion of Blocked Polyurethane Composites. *J. Appl. Polym. Sci.* **1993**, *50*, 759.

(24) Ma, C. C.; Lee, M., K. Y.; Hwang, J. S. The Correlations of Processing Variables for Optimizing the Pultrusion Process. *SAMPE J*. **1986**, 42.

(25) Yn, M. S.; Ma, C. C.; Lin, S. H. Pultrusion of Poly(ϵ-caprolactam)/Poly(butadiene-*co*-arcylonitrile) Composites: I. Simulation and a Mathematical Model. *Compos. Sci. Technol.* **1995**, *54*, 121.

(26) Chen, C. H.; Ma, C. C. M. Pultruded Fiber-reinforced PMMA/PU IPN Composites: Processability and Mechanical Properties. *Composites, Part A* **1997**, *28*, 65.

(27) Montgomery, D. C. *Design and Analysis of Experiments*; Wiley: New York, 1997.

(28) Peace; Stuart, G. *Taguchi Method: A Hands-on Approach*; Addison-Wesley: Reading, MA, 1994.

(29) Shibata, R. A Symptotically Efficient Selection of the Order of the Model for Estimating Parameter of a Linear Process. *Ann. Stat*. **1980**, *8*, 147.

(30) Grossman, T.; Lapedes, A. Noise Sensitivity Signatures for Model Selection. *Proceedings of the 12th IAPR International Conference on Pattern Recognition, Conference B: Computer Vision and Image Processing*; IEEE: New York, 1994; Vol. 2, p 213.

(31) Akaike, H. A New Look at the Statistical Model Identification. *IEEE Trans. Automat. Control* **1974**, *19*, 716.

(32) Taguchi, G. *System of Experimental Design*; Kraus International Publications: New York, 1987.

(33) Han, C. D.; Lee, D. S.; Chin, H. B. Development of a Mathematical Model for the Pultrusion Process. *Polym. Eng. Sci.* **1986**, *26*, 393.

(34) Batch, G. L.; Macosko, C. W. Heat Transfer and Cure in Pultrusion: Model and Experimental Verification. *AIChE J*. **1993**, *39*, 1228.

(35) Hall, P.; Heyde, C. C. *Martingale Limit Theory and Its Applications*; Academic Press: New York, 1980.

Received for review June 1, 2000
Revised manuscript received May 10, 2001
Accepted June 1, 2001

IE000543H