

## Research Article

# Multiple Model Identification for a High Purity Distillation Column Process Based on EM Algorithm

**Weili Xiong, Lei Chen, Fei Liu, and Baoguo Xu**

*Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China*

Correspondence should be addressed to Weili Xiong; [greenpre@163.com](mailto:greenpre@163.com)

Received 23 October 2013; Accepted 3 January 2014; Published 13 February 2014

Academic Editor: Shuping He

Copyright © 2014 Weili Xiong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Due to the strong nonlinearity and transition dynamics between different operating points of the high purity distillation column process, it is difficult to use a single model for modeling such a process. Therefore, the multiple model based approach is introduced for modeling the high purity distillation column plant under the framework of the expectation maximization (EM) algorithm. In this paper, autoregressive exogenous (ARX) models are adopted to construct the local models of this chemical process at different operating points, and the EM algorithm is used for identification of local models as well as the probability that each local model takes effect. The global model is obtained by aggregating the local models using an exponential weighting function. Finally, the simulation performed on the high purity distillation column demonstrates the effectiveness of the proposed method.

## 1. Introduction

The increasing complexity of the process control systems poses a great challenge for process modeling and parameter identification and control. In practice, most of the industrial processes show a strong nonlinearity and a dynamic nature. Therefore, it is of interest for both academic researchers and industrial practitioners to study how to model these systems and to achieve satisfactory or even optimal control performance. Linear modeling technologies have been quite sophisticated after decades of development; however, due to the inborn nonlinearity of some production processes, the performance of single linear model based controllers or optimizers may be compromised or even unsatisfactory [1]. To overcome the limitations imposed by the nonlinearity of the process, various strategies have been developed for nonlinear process modeling. Some of the popular modeling approaches have been proposed and applied in industrial processes, such as artificial neural networks [2, 3], support vector regression and its extension least squares support vector regression [4], Gaussian process regression [5], and hybrid methods [6]. The obvious feature of these methods is that they do not need any prior knowledge about the system. However, these methods do not always capture the process

dynamics well and may lose validity when the system transits among a large range of operation.

The multiple model approach is a good compromise which produces complex models or controllers by piecing a number of simpler subsystems together. Some multiple modeling strategies and applications have been investigated because they are able to represent complex nonlinear processes using the linearization technique [7]. The terminology of linear parameter varying (LPV) modeling method, which is featured by its linear structure and varying model parameters, was introduced by Shamma and Athans [8, 9]. The work on identification of multimodel LPV models can be found in much literature. For example, Zhu and Xu proposed an LPV model based on blended linear models and added weights at the input side [10, 11]. The identification of LPV models in an input-output setting with Box-Jenkins (BJ) model structure was addressed by Zhao et al. and prediction error method was employed for parameter interpolation [12]. Huang et al. studied the work of the multiple LPV model identification using several weighting functions with two scheduling variables [13]. The multiple model based recursive least square parameter identification was illustrated by Chen and Liu [14]. Jin et al. proposed a multiple model LPV approach for nonlinear systems under the framework of

the expectation maximization (EM) algorithm [15]. It is an important step on LPV identification in which the model identity is treated as the missing variable in the EM scheme. Based on the EM algorithm, the local models are synthesized with process data to obtain a global model for approximating a nonlinear process. Deng and Huang extended the identification method to nonlinear parameter varying systems in which the local model takes the form of state space equation. The identification problem is solved by using the particle filter based on EM algorithm [16]. The efficiency is illustrated on numerical examples and a pilot-scale hybrid water tank. Chen et al. continued to study the uncertain scheduling variable problem in multiple model based nonlinear identification. Both the linear and the nonlinear dynamics of scheduling variables are taken into account [17].

The distillation column is considered as one of the most important unit operations in chemical engineering [18, 19]. It becomes a benchmark problem in chemical process control for nonlinear modeling technology. The difficulties in modeling and controlling distillation columns lie in their highly nonlinear characteristics. Modeling of distillation columns may be classified into three groups: fundamental modeling, empirical modeling, and hybrid modeling [18]. In this paper, a hybrid solution based on the multiple models is adopted, which combines the two main advantages of linear and nonlinear models. The distillation column model is a nonlinear global model but can be represented as a reunion of linear models, one linear model for each operating point (local model or submodel), where each local model works at different operating point and can be identified through the iteration procedure of the EM algorithm. Finally, the global nonlinear model can be obtained by using the exponential function which takes full account of each operating point.

The remainder of this paper is organized as follows: the high purity distillation column used to test the multiple model method is described in Section 2. Section 3 lays out the mathematical formulation of the multiple model approach in which the ARX model is used as the local model. In Section 4, the identification procedure under the EM algorithm is provided. The simulation on the high purity distillation column with multiple inputs is given in Section 5. Finally, conclusions are given in Section 6.

## 2. Description of a High Purity Distillation Column

Distillation is simply defined as a process in which a liquid or vapor mixture of two or more substances is separated into its component fractions of desired purity, by the application and removal of heat [20]. The process is based on the fact that the vapor of a boiling mixture will be richer in the components that have lower boiling points. In most cases, the distillation is operated at a continuous steady state. New feed is always being added to the distillation column and products are always being removed. A high purity distillation column is strongly nonlinear, and any realistic study should take this into account [21, 22]. Distillation column has become a favorite subject in the process control engineering field, including the areas of soft-sensor, process modeling, and

TABLE 1: Notation of the distillation column in reflux and boil-up construction.

Variable	Description	Unit
1 $F$	Feed rate	kmol/min
2 $yD$	Distillate top product composition	mole fraction
3 $xB$	Distillate bottom product composition	mole fraction
4 $LT$	Reflux flow	kmol/min
5 $VB$	Boil-up flow	kmol/min

control. There are a variety of configurations for distillation columns, and each performs specific types of separations [23].

A typical two-product distillation column on a pilot plant scale is shown in Figure 1. Some important notation is summarized in Table 1. It is a nonlinear model of a distillation column with  $NT - 1$  theoretical stages including a reboiler plus a total condenser [24]. In the model, the following assumptions are made:

- (1) two components (binary separation);
- (2) one feed and two products;
- (3) constant relative volatility;
- (4) constant molar flows (same vapor flow on all stages);
- (5) no vapor holdup;
- (6) total condenser.

## 3. Problem Formulation Using Multiple Model Method

The high purity distillation column as mentioned above is a classic chemical production process; the plant may transit among several operating points. A bank of ARX models is employed to approximate the nonlinear dynamic process around each operating point. Consider the following ARX model [15, 17]:

$$y_k = \theta_{I_k}^T x_k + e_k, \quad (1)$$

where  $\{y_k \in \mathbb{R}, k = 1, 2, \dots, N\}$  is the output.  $e_k$  is the Gaussian noise with zero mean and variance  $\sigma$  added to the ARX model.  $x_k \in \mathbb{R}^n$  is the regressor vector,

$$x_k \triangleq [y_{k-1}, y_{k-2}, \dots, y_{k-n_a}, u_{k-1}^T, u_{k-2}^T, \dots, u_{k-n_b}^T]^T, \quad (2)$$

where  $\{u_k \in \mathbb{R}^m, k = 1, 2, \dots, N\}$  is the input of the nonlinear system,  $n_a$  and  $n_b$  are the orders of the output and input, and  $n = n_a + m \cdot n_b$ . Many industrial processes are often operated in certain "orderly" ways to meet different production objectives. Such orderly ways are also referred to as operating trajectory consisting of several predesigned operating points which are defined as the scheduling variable,  $W = w_{1:N} = \{w_1, w_2, \dots, w_N\}$ . Here we assume that  $M$  different operating points are known a priori, corresponding to a certain  $\{T_i, i = 1, 2, \dots, M\}$ , namely,  $T_{1:M}$ .

Around the small region of each of the operation points, the ARX local model can be used to approximate the process

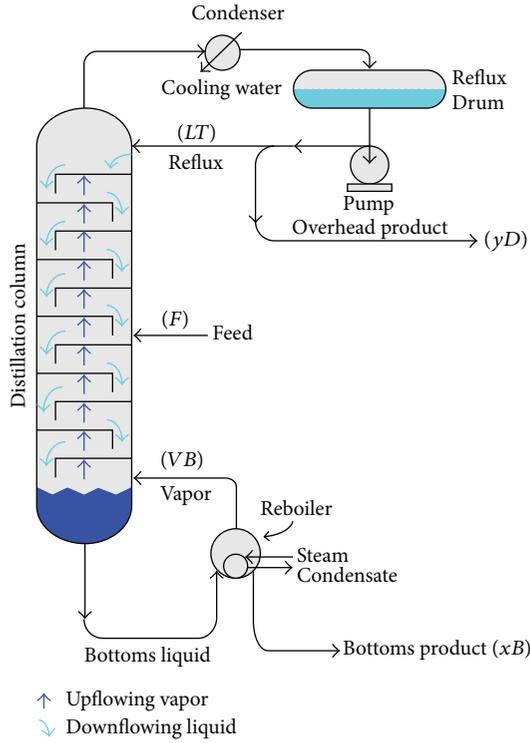


FIGURE 1: The two-product distillation column in reflux and boil-up construction.

dynamics. Here parameter  $\theta_i$  is used to represent the  $i$ th local model parameters. Let  $U = u_{1:N}$  denote the sequence of input data set  $\{u_1, u_2, \dots, u_N\}$  and let  $Y = y_{1:N}$  denote the observed output data set  $\{y_1, y_2, \dots, y_N\}$ . Here a hidden variable  $I = I_{1:N} = \{I_1, I_2, \dots, I_N\}$  is introduced to stand for the identity of each local model that takes effect at sampling time point  $k$ . So the parameters to be estimated for each local model are  $\Theta_{I_k} = \{\theta_{I_k}, \sigma\}$ . Given all of the past observed data information, the probability of the output  $y_k$  can be computed as

$$\begin{aligned}
 P(y_k | y_{k-1:1}, u_{k:1}, w_{k:1}, T_{M:1}, \Theta) \\
 &= \sum_{i=1}^M p(y_k, I_k = i | y_{k-1:1}, u_{k:1}, w_{k:1}, T_{M:1}, \Theta) \\
 &= \sum_{i=1}^M p(y_k | I_k = i, y_{k-1:1}, u_{k:1}, w_{k:1}, T_{M:1}, \Theta) \\
 &\quad \times p(I_k = i | y_{k-1:1}, u_{k:1}, w_{k:1}, T_{M:1}, \Theta),
 \end{aligned} \tag{3}$$

Based on the fact that the observed output  $y_k$  at  $k$ th sampling time instant directly depends on the regressor vector  $x_k$  as well as the data identity of  $I_k$ , (3) can be further written as

$$\sum_{i=1}^M p(y_k | I_k = i, y_{k-1:k-n_a}, y_{k-n_a-1:1}, u_k, u_{k-1:k-n_b}, u_{k-n_b-1:1}, w_{k:1}, T_{M:1}, \Theta),$$

$$\begin{aligned}
 p(I_k = i | y_{k-1:1}, u_{k:1}, w_{k:1}, T_{M:1}, \Theta) \\
 &= \sum_{i=1}^M p(y_k | x_k, I_k = i, \Theta) p(I_k = i | w_k, T_{M:1}) \\
 &= \sum_{i=1}^M p(y_k | x_k, I_k = i, \Theta) \cdot \alpha_{k,i}.
 \end{aligned} \tag{4}$$

Here  $\alpha_{k,i}$  is the normalized weight standing for the probability that the  $i$ th submodel takes effect at time  $k$ . An exponential function is adopted as

$$\begin{aligned}
 \alpha_{k,i} &= p(I_k = i | w_k, T_{M:1}) \\
 &= \frac{\exp(-(w_k - T_i)^2 / 2(o_i)^2)}{\sum_{i=1}^M \exp(-(w_k - T_i)^2 / 2(o_i)^2)},
 \end{aligned} \tag{5}$$

where  $w_k$  is the scheduling variable at the sampling time  $k$  and  $o_i$  is defined as the validity width of the  $i$ th local model which is limited between  $o_{\min}$  (the lower bound for  $o_i, i = 1, 2, \dots, M$ ) and  $o_{\max}$  (the upper bound for  $o_i, i = 1, 2, \dots, M$ ).

The missing data set is the random variable  $I = \{I_1, I_2, \dots, I_N\}$  denoting the identity of each local model, namely,  $C_{\text{mis}} = \{I\}$ . The values of input variables, output variables, and the scheduling variable are observed or given a priori, so  $C_{\text{obs}} = \{U, Y, W, T_{1:M}\}$ . The parameters to be estimated from process input and output data under the framework of the EM algorithm are  $\Theta = \{\theta_i, o_i, \sigma\}_{i=1,2,\dots,M}$ .

## 4. The ARX Model Identification under the EM Algorithm

**4.1. The EM Algorithm Review.** The EM algorithm is an efficient iterative procedure for maximum likelihood estimation in the presence of missing data. The main strength of the EM algorithm lies in its ability to handle missing variables in the identification data set [25]. After being first introduced by Raghavan et al., it has found extensive applications in various areas for finding maximum likelihood estimates of parameters in probabilistic models [26]. In the EM procedure, both the complete data log likelihood  $p(C_{\text{obs}}, C_{\text{mis}} | \Theta)$  and the conditional predictive distribution of missing data  $p(C_{\text{mis}} | C_{\text{obs}}, \Theta^{\text{old}})$  are needed to be calculated. There are two steps, namely, the expectation step ( $E$ -step) and the maximization step ( $M$ -step). The EM algorithm proceeds as follows.

- (1) Initialization: give an initial value of the model parameter vector.
- (2)  $E$ -step: the  $Q$ -function is calculated, given the current parameter  $\Theta$  estimated from the previous iteration,

$$Q(\Theta | \Theta^{\text{old}}) = E_{C_{\text{mis}} | (C_{\text{obs}}, \Theta^{\text{old}})} \{ \log p(C_{\text{obs}}, C_{\text{mis}} | \Theta) \}. \tag{6}$$

- (3)  $M$ -step: maximize the  $Q$ -function with respect to  $\Theta$  to obtain the new iterative parameter estimate.

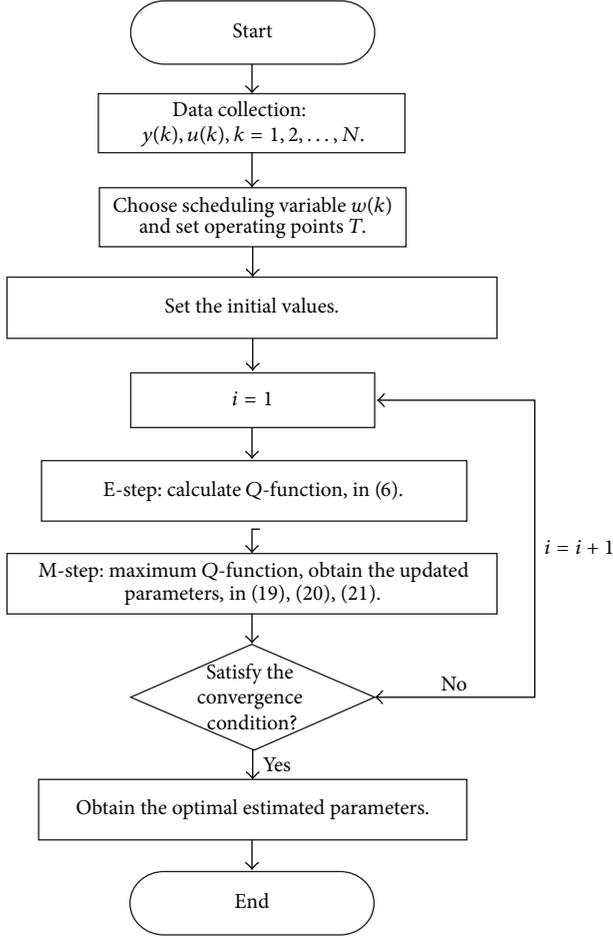


FIGURE 2: The flowchart of the EM algorithm for estimating the parameters.

- (4) Iteration: step 2 and step 3 are repeated until the changes in parameters after each iteration are within a specified tolerance level.

The above steps ensure that the log likelihood function of the observed data increases at every step. So the EM algorithm is guaranteed to converge to some maximum of the likelihood function [15]. The convergence performance under general conditions is given in Wu [27]. The flowchart of the algorithm is shown in Figure 2.

**4.2. Formulation of the Multiple Model Approach.** As mentioned above, if we compute the Q-function,  $p(C_{\text{obs}}, C_{\text{mis}} | \Theta)$  should be derived at first using the probability chain rule:

$$\begin{aligned}
 p(C_{\text{obs}}, C_{\text{mis}} | \Theta) &= p(Y, U, W, I, T_{1:M} | \Theta) \\
 &= p(Y | U, W, I, T_{1:M}, \Theta) p(I | U, W, T_{1:M}, \Theta) \\
 &\quad \times p(W | U, T_{1:M}, \Theta) p(U, T_{1:M} | \Theta).
 \end{aligned} \tag{7}$$

Using the Bayesian theory, the joint probability density function such as the first term of (7) can be written as

$$\begin{aligned}
 p(Y | U, W, I, T_{1:M}, \Theta) &= p(y_{1:N} | u_{1:N}, w_{1:N}, I_{1:M}, T_{1:M}, \Theta) \\
 &= p(y_N | y_{N-1:1}, u_{1:N}, w_{1:N}, I_{1:N}, T_{1:M}, \Theta) \\
 &\quad \times p(y_{N-1} | y_{N-2:1}, u_{1:N}, w_{1:N}, I_{1:N}, T_{1:M}, \Theta) \\
 &\quad \cdots p(y_2 | y_1, u_{1:N}, w_{1:N}, I_{1:N}, T_{1:M}, \Theta) \\
 &\quad \times p(y_1 | u_{1:N}, w_{1:N}, I_{1:N}, T_{1:M}, \Theta) \\
 &= \prod_{k=1}^N p(y_k | x_k, w_{1:N}, I_{1:N}, T_{1:M}, \Theta) \\
 &= \prod_{k=1}^N p(y_k | x_k, \Theta_{I_k}).
 \end{aligned} \tag{8}$$

Similarly, the model identity  $I_k$  can be determined from (5), so the second term of (7) can be simplified as

$$\begin{aligned}
 p(I | U, W, T_{1:M}, \Theta) &= p(I_{1:N} | u_{1:N}, w_{1:N}, T_{1:M}, \Theta) \\
 &= p(I_N | y_{N-1:1}, u_{1:N}, w_{1:N}, T_{1:M}, \Theta) \\
 &\quad \times p(I_{N-1} | I_{N-2:1}, u_{1:N}, w_{1:N}, T_{1:M}, \Theta) \\
 &\quad \cdots p(I_2 | I_1, u_{1:N}, w_{1:N}, T_{1:M}, \Theta) \\
 &\quad \times p(I_1 | u_{1:N}, w_{1:N}, T_{1:M}, \Theta) \\
 &= \prod_{k=1}^N p(I_k | I_{k-1:1}, u_{1:N}, w_{1:N}, T_{1:M}, \Theta) \\
 &= \prod_{k=1}^N p(I_k | w_k, T_{1:M}, \Theta).
 \end{aligned} \tag{9}$$

The derivation of the third term in (7) is based on the chain rule and the assumption that the scheduling variable  $w_k$  is known a priori and independent of the past observations as well as the model parameters. Based on the fact that most of the chemical plants predesign the operating points, namely, scheduling variable,  $p(W | U, T_{1:M}, \Theta)$  is derived as

$$\begin{aligned}
 p(W | U, T_{1:M}, \Theta) &= p(w_N | w_{N-1:1}, u_{1:N}, T_{1:M}, \Theta) \\
 &\quad \times p(w_{N-1} | w_{N-2:1}, u_{1:N}, T_{1:M}, \Theta) \\
 &\quad \cdots p(w_2 | w_1, u_{1:N}, T_{1:M}, \Theta) \\
 &\quad \times p(w_1 | u_{1:N}, T_{1:M}, \Theta)
 \end{aligned}$$

$$\begin{aligned}
 &= \prod_{k=1}^N p(w_k | w_{k-1:1}, u_{1:N}, T_{1:M}, \Theta) \\
 &= \prod_{k=1}^N p(w_k).
 \end{aligned} \tag{10}$$

Then we can substitute (8), (9), and (10) into (7), and the joint probability density of the likelihood of the full data set can be rewritten as

$$\begin{aligned}
 &p(C_{\text{obs}}, C_{\text{mis}} | \Theta) \\
 &= p(Y, U, W, I, T_{1:M} | \Theta) \\
 &= \prod_{k=1}^N p(y_k | x_k, \Theta_{I_k}) \prod_{k=1}^N p(I_k | w_k, T_{1:M}, o) \\
 &\quad \times \prod_{k=1}^N p(w_k) p(U, T_{1:M} | \Theta) \\
 &= \prod_{k=1}^N p(y_k | x_k, \Theta_{I_k}) \prod_{k=1}^N p(I_k | w_k, T_{1:M}, o) C_0,
 \end{aligned} \tag{11}$$

where

$$\begin{aligned}
 &\prod_{k=1}^N p(w_k) = C_1, \\
 &p(U, T_{1:M} | \Theta) = C_2, \\
 &C_0 = C_1 \cdot C_2.
 \end{aligned} \tag{12}$$

The expectation of the complete data in the procedure of the EM algorithm is

$$\begin{aligned}
 &Q(\Theta | \Theta^{\text{old}}) \\
 &= E_{I|(\Theta^{\text{old}}, C_{\text{obs}})} \log \left\{ \prod_{k=1}^N p(y_k | x_k, \Theta_{I_k}) \right. \\
 &\quad \left. \times \prod_{k=1}^N p(I_k | w_k, T_{1:M}, o) C_0 \right\} \\
 &= E_{I|(\Theta^{\text{old}}, C_{\text{obs}})} \left\{ \sum_{k=1}^N \log p(y_k | x_k, \Theta_{I_k}) \right. \\
 &\quad \left. + \sum_{k=1}^N \log p(I_k | w_k, T_{1:M}, o) + \log C_0 \right\} \\
 &= \sum_{I_1} \sum_{I_2} \cdots \sum_{I_N} \left\{ \sum_{k=1}^N \log p(y_k | x_k, \Theta_{I_k}) \right. \\
 &\quad \left. + \sum_{k=1}^N \log p(I_k | w_k, T_{1:M}, o) + \log C_0 \right\} \\
 &\quad \cdot p(I | \Theta^{\text{old}}, C_{\text{obs}})
 \end{aligned}$$

$$\begin{aligned}
 &= \sum_{i=1}^M \left\{ \sum_{k=1}^N \log p(y_k | x_k, \Theta_i) \right. \\
 &\quad \left. + \sum_{k=1}^N \log p(I_k = i | w_k, T_{1:M}, o_i) + \log C_0 \right\} \\
 &\quad \cdot p(I_k = i | \Theta_i^{\text{old}}, C_{\text{obs}}) \\
 &= \sum_{k=1}^N \sum_{i=1}^M \log p(y_k | x_k, \Theta_i) p(I_k = i | \Theta_i^{\text{old}}, C_{\text{obs}}) \\
 &\quad + \sum_{k=1}^N \sum_{i=1}^M \log p(I_k = i | w_k, T_{1:M}, o_i) \\
 &\quad \quad \times p(I_k = i | \Theta_i^{\text{old}}, C_{\text{obs}}) \\
 &\quad + \sum_{k=1}^N \sum_{i=1}^M \log C_0 p(I_k = i | \Theta_i^{\text{old}}, C_{\text{obs}}).
 \end{aligned} \tag{13}$$

From (13) we need to compute the following probability density function:

- (1)  $\log p(y_k | x_k, \Theta_i)$ ,
- (2)  $p(I_k = i | w_k, T_{1:M}, o_i)$ ,
- (3)  $p(I_k = i | \Theta_i^{\text{old}}, C_{\text{obs}})$ ,

where  $p(y_k | x_k, \Theta_i)$  is calculated by

$$p(y_k | x_k, \Theta_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp \frac{-1}{2\sigma^2} (y_k - \theta_i^T x_k)^T (y_k - \theta_i^T x_k), \tag{14}$$

$$\begin{aligned}
 &\log p(y_k | x_k, \Theta_i) \\
 &= -\log \sqrt{2\pi}\sigma + \frac{-1}{2\sigma^2} (y_k - \theta_i^T x_k)^T (y_k - \theta_i^T x_k).
 \end{aligned} \tag{15}$$

$p(I_k = i | w_k, T_{1:M}, o_i)$  is given by (5).

The last probability density function to be calculated denotes the conditional probability that the  $k$ th data point comes from the  $i$ th local model and can be derived from the Bayesian rule:

$$\begin{aligned}
 &p(I_k = i | \Theta_i^{\text{old}}, C_{\text{obs}}) \\
 &= p(I_k = i | \Theta_i^{\text{old}}, y_{1:N}, u_{1:N}, w_{1:N}, T_{1:M}) \\
 &= p(y_{1:N}, u_{1:N}, w_{1:N}, T_{1:M}, \Theta_i^{\text{old}} | I_k = i) \cdot p(I_k = i) \\
 &\quad \times \left( \sum_{i=1}^M p(y_{1:N}, u_{1:N}, w_{1:N}, T_{1:M}, \Theta_i^{\text{old}} | I_k = i) \right. \\
 &\quad \left. \times p(I_k = i) \right)^{-1}
 \end{aligned}$$

$$\begin{aligned}
&= p(y_N | y_{1:N-1}, u_{1:N}, w_{1:N}, T_{1:M}, \Theta_i^{\text{old}}, I_k = i) \\
&\quad \cdot p(I_k = i, y_{1:N-1}, u_{1:N}, w_{1:N}, T_{1:M}, \Theta_i^{\text{old}}) \\
&\quad \times \left( \sum_{i=1}^M p(y_N | y_{1:N-1}, u_{1:N}, w_{1:N}, T_{1:M}, \Theta_i^{\text{old}}, I_k = i) \right. \\
&\quad \quad \left. \times p(I_k = i, y_{1:N-1}, u_{1:N}, w_{1:N}, T_{1:M}, \Theta_i^{\text{old}}) \right)^{-1} \\
&= p(y_k | x_k, \Theta_i^{\text{old}}) p(I_k = i | w_k, T_{1:M}, o_i) \\
&\quad \times \left( \sum_{i=1}^M p(y_k | x_k, \Theta_i^{\text{old}}) p(I_k = i | w_k, T_{1:M}, o_i) \right)^{-1}. \tag{16}
\end{aligned}$$

4.3. *The Summary of the Identification Procedure.* Now we give the identification procedure under the frame of the EM algorithm as follows.

(1) *Initialization.* Set the initial value of parameters to be estimated.

(2) *E-step.* Using the current parameter  $\Theta^{\text{old}}$ , calculate  $p(I_k = i | (\Theta_i^{\text{old}}, C_{\text{obs}}))$ , where  $p(I_k = i | w_k, T_{1:M}, o_i)$  can be computed by (5) and  $\log p(y_k | x_k, \Theta_i)$  is calculated from (15).

(3) *M-step.* Based on the three terms of the function  $Q(\Theta | \Theta^{\text{old}})$ , only the term as (15) depends on the local model parameter  $\theta_i$  and the process noise variance  $\sigma^2$ . Derivative is taken over this term with respect to  $\theta_i$  and  $\sigma^2$ , respectively. So the new parameter estimation can be calculated as follows:

$$\begin{aligned}
\frac{\partial Q(\Theta | \Theta^{\text{old}})}{\partial \theta_i} &= \frac{\partial}{\partial \theta_i} \sum_{k=1}^N \sum_{i=1}^M \log p(y_k | x_k, \Theta_i) \\
&\quad \times p(I_k = i | \Theta_i^{\text{old}}, C_{\text{obs}}) = 0. \\
\frac{\partial Q(\Theta | \Theta^{\text{old}})}{\partial \sigma^2} &= \frac{\partial}{\partial \sigma^2} \sum_{k=1}^N \sum_{i=1}^M \log p(y_k | x_k, \Theta_i) \\
&\quad \times p(I_k = i | \Theta_i^{\text{old}}, C_{\text{obs}}) = 0. \tag{17}
\end{aligned}$$

As a result, the new  $\theta_i$  of the  $i$ th ARX local model and  $\sigma^2$  are equal to

$$\theta_i^{\text{new}} = \frac{\sum_{k=1}^N p(I_k = i | \Theta_i^{\text{old}}, C_{\text{obs}}) x_k^T y_k}{\sum_{k=1}^N p(I_k = i | \Theta_i^{\text{old}}, C_{\text{obs}}) x_k^T x_k}, \tag{18}$$

$$\begin{aligned}
\sigma_{\text{new}}^2 &= \sum_{k=1}^N \sum_{i=1}^M p(I_k = i | \Theta_i^{\text{old}}, C_{\text{obs}}) \\
&\quad \times (y_k - (\theta_i^{\text{new}})^T x_k)^T (y_k - (\theta_i^{\text{new}})^T x_k) \\
&\quad \times \left( \sum_{k=1}^N \sum_{i=1}^M p(I_k = i | \Theta_i^{\text{old}}, C_{\text{obs}}) \right)^{-1}. \tag{19}
\end{aligned}$$

For the local model validity width  $o_i$ ,  $i = 1, 2, \dots, M$ , due to the usage of exponential function, it cannot be calculated from an analytical form directly. The optimization problem searching for the optimal value of  $o_i$  can be expressed as

$$\begin{aligned}
\max_{o_i, i=1,2,\dots,M} &\sum_{k=1}^N \sum_{i=1}^M p(I_k = i | (\Theta_i^{\text{old}}, C_{\text{obs}})) \\
&\quad \times \log p(I_k = i | w_k, T_{1:M}, o_i) \tag{20}
\end{aligned}$$

$$\text{s.t. } o_{\min} \leq o_i \leq o_{\max}, \quad i = 1, 2, \dots, M. \tag{21}$$

Here a nonlinear numerical optimization is required to obtain the optimal  $o_i$  between the  $o_{\min}$  and  $o_{\max}$ . A constrained nonlinear optimization function named “fmincon” in optimization toolbox of “Matlab” is adopted to search for the optimal  $o_i$  at each of the iteration steps [15].

(4) *Iteration.* Repeat step (2) and step (3), until the converge condition is satisfied. Here the converge condition is defined as the changes of the estimated value between two iterations which are as small as the level  $10^{-6}$ .

## 5. A Distillation Column Simulation

Nonlinear system identification is much more complex than linear system identification. In this section, a high purity distillation column is used to test the performance of the multiple model approach based on the EM algorithm.

It has been shown that the purer the products get, the more nonlinear the system becomes [20]. A distillation column is also a typical example for a MISO system in which there are strong interactions between the variables. In this paper, choosing LV-configuration is motivated by the fact that the LV-configuration is the most commonly used in industrial practice.

Here, the process input variables available for control purpose are the reflux ( $LT$ ) and the boil-up rate ( $VB$ ). The top product composition ( $yD$ ) variable is the output. The disturbances to a distillation column can come from many sources. Besides the two manipulated variables, they can come from the feed (feed rate and feed composition), from the pressure inside the column and from the cooling water, and so forth. The distillation column is very sensitive to the feed rate disturbance. The feed rate is assumed to be measurable; therefore, its effect can be incorporated in the EM algorithm. In this paper, the feed rate  $F$  is treated as the scheduling variable [23]. When the column is operated over a relatively wide operating region instead of at a single set point, it reveals a significant nonlinear behaviour.

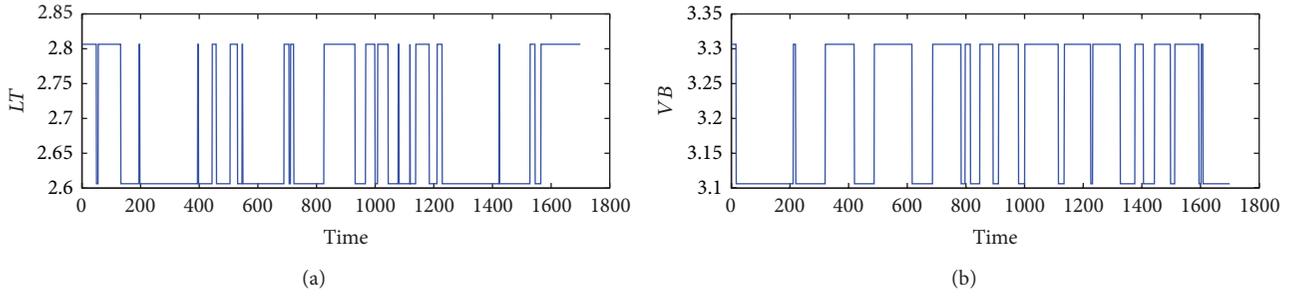


FIGURE 3: The observed input data of the distillation column.

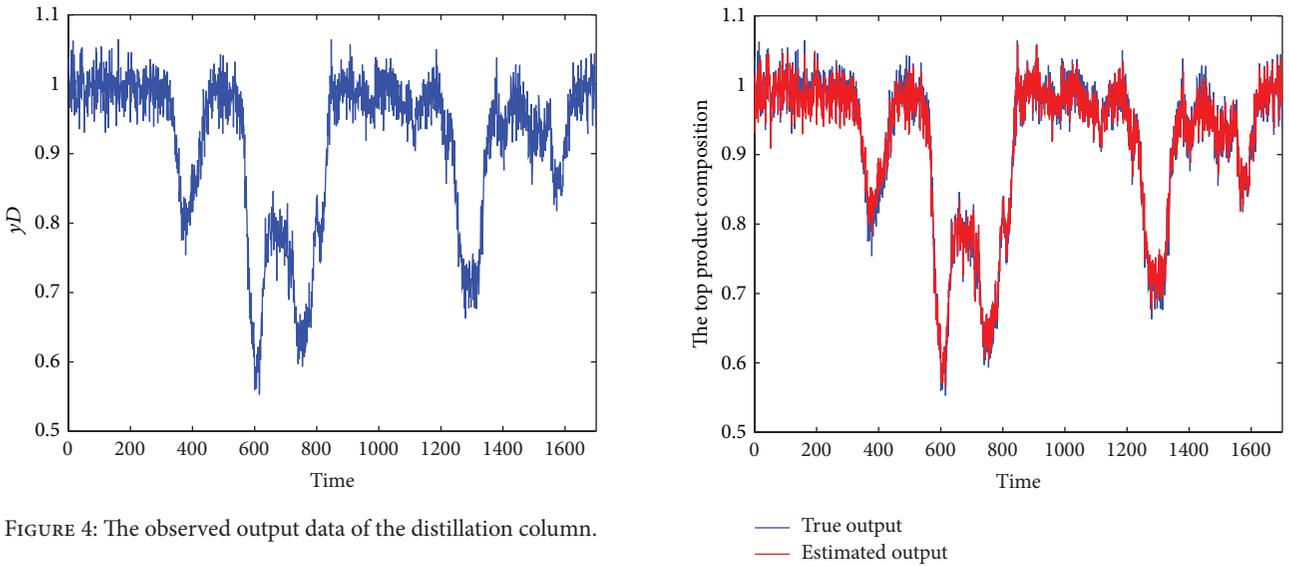


FIGURE 4: The observed output data of the distillation column.

FIGURE 6: The comparison of the distillation column model.

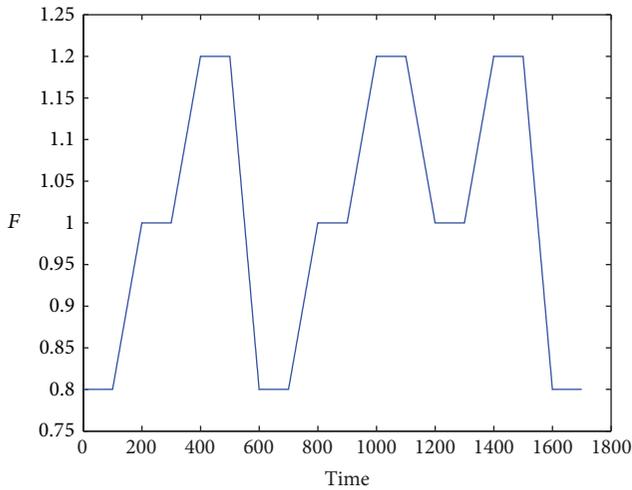


FIGURE 5: The scheduling variable of the distillation column.

In the simulation, white noise with a variance of 0.05 is added in the simulated output data for unknown process noise. The observed input and output data for the high purity distillation column are shown in Figures 3 and 4. The scheduling variable is designed with three operating points at 0.8 kmol/min, 1.0 kmol/min, and 1.2 kmol/min, respectively,

which were given in Figure 5. In order to achieve high identification performance the random binary excitation signals with small magnitude are added in the input variables. The feed rate increases by a fixed step size and no additional excitation signal is added. Then applying the multiple model modeling approach and using the EM algorithm to estimate the parameters of each local model, the simulation results of self-validation tests indicate the efficiency of the proposed method. The comparison and the weighting function curve are given in Figures 6 and 7.

To further test the performance of the identified multiple model, the cross-validation is also carried on the case. The data used for cross-validation as shown in Figures 8 and 9 are generated under different feed rates from Figure 5. Two different operating points are selected in the cross-validation simulation as in Figure 10. It can be seen from Figure 11 that the real process data and the model prediction are in good agreement with each other which effectively confirms the efficiency of the proposed algorithm.

## 6. Conclusions

A multiple model modeling algorithm based on EM algorithm is proposed and applied for a high purity distillation

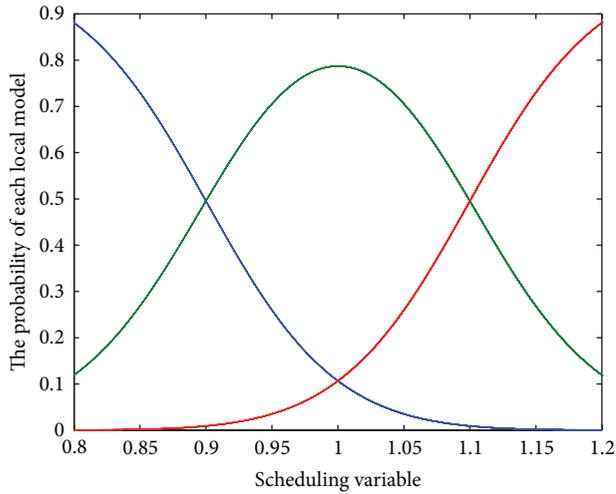


FIGURE 7: The weight of each local model at different operating points.

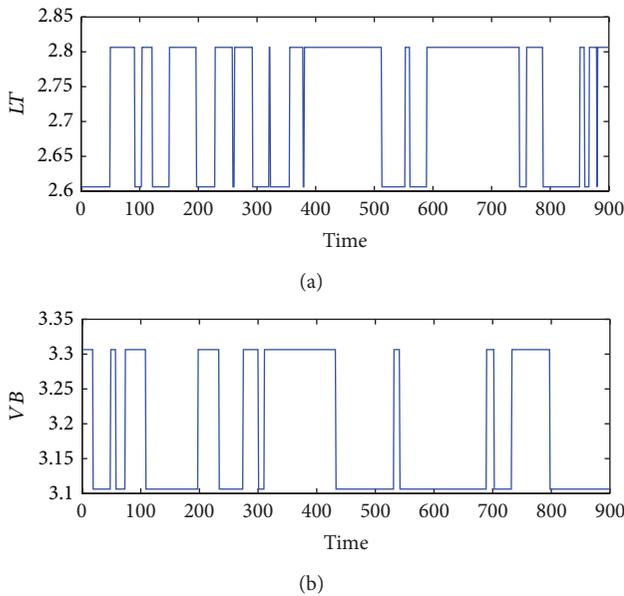


FIGURE 8: The observed input data of the distillation column in cross-validation simulation.

column reaction process. Taking a full consideration of the dynamic, nonlinearity, and complexity, the parameters of each local model are identified under different operating points. The global model can be obtained by using a weighting function, which adopts an exponential form about scheduling variable under the framework of EM algorithm. Simulation results show that the fusion model of multiple models based on the EM algorithm is very effective in the chemical process of a high purity distillation column.

### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

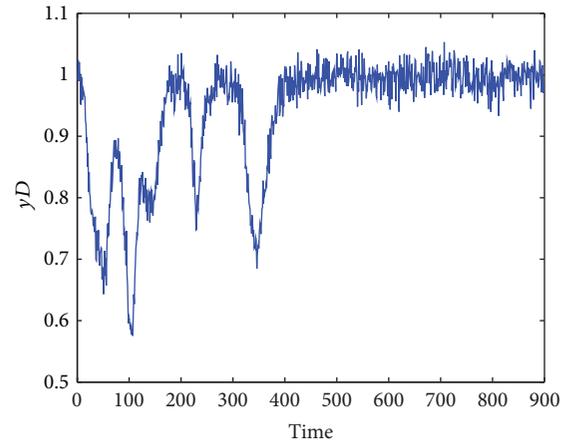


FIGURE 9: The observed output data of the distillation column in cross-validation simulation.

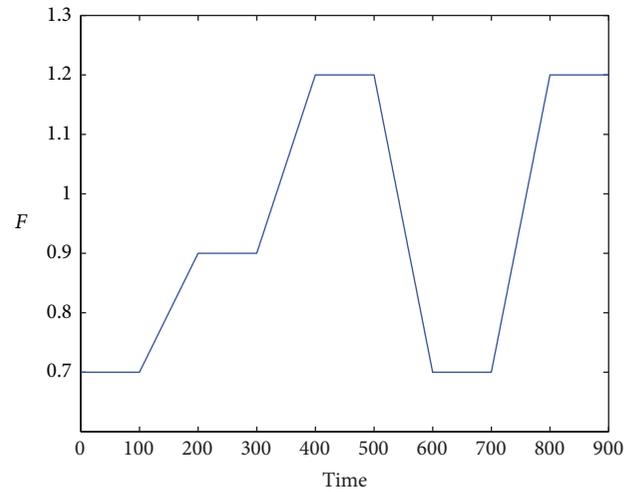


FIGURE 10: The scheduling variable of the distillation column in cross-validation simulation.

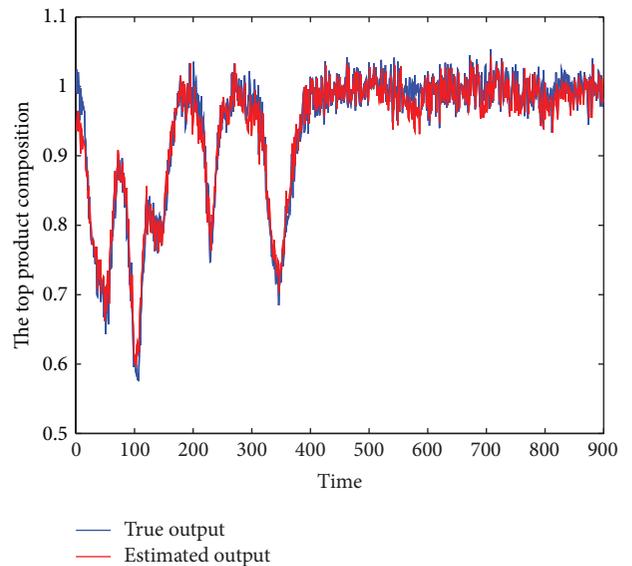


FIGURE 11: The comparison of the distillation column model in cross-validation simulation.

## Acknowledgments

The authors thank the financial support by the National Natural Science Foundation of China (nos.: 21206053, 21276111, and 61273131) and the paper is partly supported by the 111 Project (B12018), the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD), the Fundamental Research Funds for the Central Universities (JUDCF10064), and Jiangsu Innovation Program for Graduates (CXZZ11\_0464).

## References

- [1] L. Ljung, *System Identification: Theory for the User*, Prentice Hall, 1987.
- [2] A. J. de Assis and R. Maciel Filho, "Soft sensors development for on-line bioreactor state estimation," *Computers and Chemical Engineering*, vol. 24, no. 2–7, pp. 1099–1103, 2000.
- [3] J. Glassey, M. Ignova, A. C. Ward, G. A. Montague, and A. J. Morris, "Bioprocess supervision: neural networks and knowledge based systems," *Journal of Biotechnology*, vol. 52, no. 3, pp. 201–205, 1997.
- [4] W. Yan, H. Shao, and X. Wang, "Soft sensing modeling based on support vector machine and Bayesian model selection," *Computers and Chemical Engineering*, vol. 28, no. 8, pp. 1489–1498, 2004.
- [5] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, Mass, USA, 2006.
- [6] X. Wang, J. Chen, C. Liu, and F. Pan, "Hybrid modeling of penicillin fermentation process based on least square support vector machine," *Chemical Engineering Research and Design*, vol. 88, no. 4, pp. 415–420, 2010.
- [7] R. M. Smith, *Multiple Model Approaches to Modeling and Control*, Taylor and Francis, London, UK, 1997.
- [8] J. S. Shamma and M. Athans, "Guaranteed properties of gain scheduled control for linear parameter-varying plants," *Automatica*, vol. 27, no. 3, pp. 559–564, 1991.
- [9] W. J. Rugh and J. S. Shamma, "Research on gain scheduling," *Automatica*, vol. 36, no. 10, pp. 1401–1425, 2000.
- [10] Y. Zhu and Z. Xu, "A method of LPV model identification for control," in *Proceedings of the 17th World Congress of the International Federation Automatic Control*, pp. 5018–5023, Seoul, Republic of Korea, 2008.
- [11] Z. Xu, J. Zhao, J. Qian, and Y. Zhu, "Nonlinear MPC using an identified LPV model," *Industrial and Engineering Chemistry Research*, vol. 48, no. 6, pp. 3043–3051, 2009.
- [12] Y. Zhao, B. Huang, H. Su, and J. Chu, "Prediction error method for identification of LPV models," *Journal of Process Control*, vol. 22, no. 1, pp. 180–193, 2012.
- [13] J. Huang, G. Ji, Y. Zhu, and P. Bosch, "Identification of multi-model LPV models with two scheduling variables," *Journal of Process Control*, vol. 22, no. 7, pp. 1198–1208, 2012.
- [14] L. Chen and F. Liu, "Recursive parameter identification for fermentation processes with the multiple model technique," *Applied Mathematical Modelling*, vol. 36, no. 5, pp. 2275–2285, 2012.
- [15] X. Jin, B. Huang, and D. S. Shook, "Multiple model LPV approach to nonlinear process identification with em algorithm," *Journal of Process Control*, vol. 21, no. 1, pp. 182–193, 2011.
- [16] J. Deng and B. Huang, "Identification of nonlinear parameter varying systems with missing output data," *American Institute of Chemical Engineers Journal*, vol. 58, no. 11, pp. 3454–3467, 2012.
- [17] L. Chen, A. Tulsyan, B. Huang, and F. Liu, "Multiple model approach to nonlinear system identification with an uncertain scheduling variable using EM algorithm," *Journal of Process Control*, vol. 23, no. 10, pp. 1480–1496, 2013.
- [18] H. S. Truong, I. Ismail, and R. Razali, "Fundamental modeling and simulation of a binary continuous distillation column," in *Proceedings of the International Conference on Intelligent and Advanced Systems (ICIAS '10)*, pp. 1–5, Kuala Lumpur, Malaysia, June 2010.
- [19] H. H. J. Bloemen, C. T. Chou, T. J. J. Van den Boom, V. Verdult, M. Verhaegen, and T. C. Backx, "Wiener model identification and predictive control for dual composition control of a distillation column," *Journal of Process Control*, vol. 11, no. 6, pp. 601–620, 2001.
- [20] A.-S. Băieșu, "Modeling a nonlinear binary distillation column," *Control Engineering and Applied Informatics*, vol. 13, no. 1, pp. 49–53, 2011.
- [21] W. L. Luyben, "Derivation of transfer functions for highly nonlinear distillation columns," *Industrial and Engineering Chemistry Research*, vol. 26, no. 12, pp. 2490–2495, 1987.
- [22] C. Fuentes and W. L. Luyben, "Control of high-purity distillation columns," *Process Design and Development*, vol. 22, no. 3, pp. 361–366, 1983.
- [23] G. Ravi Srinivas, Y. Arkun, I.-L. Chien, and B. A. Ogunnaike, "Nonlinear identification and control of a high-purity distillation column: a case study," *Journal of Process Control*, vol. 5, no. 3, pp. 149–162, 1995.
- [24] S. Skogestad, "Dynamics and control of distillation columns: a tutorial introduction," *Chemical Engineering Research and Design*, vol. 75, no. 6, pp. 539–562, 1997.
- [25] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, John Wiley & Sons, Hoboken, NJ, USA, Second edition, 2008.
- [26] H. Raghavan, A. K. Tangirala, R. B. Gopaluni, and S. L. Shah, "Identification of chemical processes with irregular output sampling," *Control Engineering Practice*, vol. 14, no. 5, pp. 467–480, 2006.
- [27] C.-F. J. Wu, "On the convergence properties of the EM algorithm," *The Annals of Statistics*, vol. 11, no. 1, pp. 95–103, 1983.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

