Korean J. Chem. Eng., **27**(5), 1366-1371 (2010)
DOI: 10.1007/s11814-010-0254-z

Context-kernel support vector machines (KSVM) based run-to-run control for nonlinear processes

Junghui Chen†, Kai-Yuan Liu, and Jose Munoz

R&D Center for Membrane Technology, Department of Chemical Engineering,
Chung-Yuan Christian University, Chung-Li, Taiwan 320, R. P. China
(*Received 4 December 2009 / accepted 17 January 2010*)

Abstract−Past studies on multi-tool and multi-product (MTMP) processes have focused on linear systems. In this paper, a novel run-to-run control (RtR) methodology designed for nonlinear semiconductor processes is presented. The proposed methodology utilizes kernel support vector machines (KSVM) to perform nonlinear modeling. In this method, the original variables are mapped using a kernel function into a feature space where linear regression is done. To eliminate the effects of unknown disturbances and drifts, the KSVM expression for the KSVM controller is modified to include constants that are updated in a manner similar to the weights used in double exponential weighting moving average method and the control law for KSVM controllers is derived. Illustrative examples are presented to demonstrate the effectiveness of KSVM and its method in process modeling and control of processes. Even if there is limited data in process modeling, KSVM still has the good capability of characterizing the nonlinear behavior. The performance of the proposed KSVM control algorithm is highly satisfactory and is superior to the other MTMP control algorithms in controlling MTMP processes.

Key words: Batch Process, Kernel Support Vector Machines, Nonlinear Processes, Run-to-run Control

INTRODUCTION

In recent years, there has been a strong interest in feedback control in the semiconductor manufacturing industry, particularly in the area of wafer processing. Many studies have been conducted to determine the right amount of control inputs at the appropriate time to get the specified process outputs. The need to get the right amount of control inputs is particularly important in reducing the incidence of producing off-spec products, and hence in lowering the cost of operation. Presently, model-based run-to-run (RtR) control, is widely used in the semiconductor industry to get the right amount of control inputs and keep quality variables close to design specifications. It adjusts the setpoints of the individual process controllers based on the ex-situ measurements of past runs and on the calculated responses of an assumed model to update the inputs for the next run [1,2]. The assumed model is usually built for a particular set of manufacturing contexts, called control threads, which specifies the combination of tools, products and other sources of product quality variations. Different batches run under the same control threads yield the same process outputs for the same control inputs.

In multi-tool and multi-product (MTMP) processes, the number of threads can be very large because of the production of many different products and the use of many tools. The production volume of different products is highly uneven. Although there are some products that have many batch runs, other products, called low-runner products, have only a few batch runs. In such a "high mix" production setting, a model built for control threads of low-runner products will not be accurate due to inadequate input-output data. This necessitates sharing of information among different control threads and building of a general model that is non-threaded. The inputs to such a model are control inputs and context variables, and the relevant parameters are tool-specific and product-specific constants [3].

The studies done on non-threaded modeling and control of MTMP processes assumed that the process model is linear. However, there are many semiconductor processes that exhibit small to large nonlinearity in behaviors. Linearization of these process models will approximate the behavior within a small range around the normal operating conditions, but the process may operate outside the range as it responds to setpoint changes and drifts due to aging of equipment. The support vector machine (SVM), a method based on statistical learning, has been used in regression and its generalization performance (i.e., error rates on test sets) either matches or is significantly better than that of the other methods. In performing linear regression, a balance is made between the accuracy of data representation and the structure of the approximating model [4,5]. By using a kernel function to map non-linearly the original variables into a higher dimensional feature space wherein linear regression is done, a modified version of SVM called kernel support vector machines (KSVM) can also do nonlinear regression [6]. KSVM is particularly suited to nonlinear modeling of MTMP processes since ordinary regression methods may not yield good results due to its small information size for each processing context.

This paper proposes a novel run-to-run control technique which utilizes KSVM for nonlinear semiconductor processes. The KSVM method used in the control algorithm is modified, which enables the KSVM control system to eliminate effects of unknown disturbances and drifts. This paper is organized into five sections. After the introduction, the second section reviews the current state of RtR control of MTMP processes and EWMA control of linear systems. The third section develops the KSVM method for modeling of nonlinear processes. Then the KSVM based controller law for control

†To whom correspondence should be addressed.
E-mail: jason@wavenet.cycu.edu.tw

of nonlinear processes is developed in Section 4. An illustrative example is given to demonstrate the effectiveness of the KSVM model and the KSVM controller in Section 5. The conclusions are given in Section 6.

REVIEW OF RUN-TO-RUN CONTROL OF MTMP PROCESSES

In the general ("non-threaded") model, the control input for batch i is designated by x_i, which is a scalar. The process output is designated by y_i, which is a scalar. Aside from control inputs, additional inputs need to be provided in terms of context variables. The context variables for batch i determine the state of processing encountered in run i and are designated by a row vector $\mathbf{a}_{o,i}$, which consists of mostly zeros and a few ones. For a multi-tool and multi-product process of K tools and D products, $\mathbf{a}_{o,i}$ would contain K+D elements. The first K elements refer to the tool category and the lth element of $\mathbf{a}_{o,i}$ is nonzero only if the lth tool is used. The last D elements refer to the product category and the K+m th element of $\mathbf{a}_{o,i}$ is nonzero if m th product is produced.

For N runs, the control input vector $\mathbf{x}$, process output matrix $\mathbf{y}$ and context matrix $\mathbf{A}_o$ are defined in Eqs. (1) to (3).

$$\mathbf{x}=[x_1\ x_2 \cdots x_N]^T \tag{1}$$

$$\mathbf{y}=[y_1\ y_2 \cdots y_N]^T \tag{2}$$

$$\mathbf{A}_o=\begin{bmatrix}\mathbf{a}_{o,1}\\ \mathbf{a}_{o,2}\\ \vdots \\ \mathbf{a}_{o,N}\end{bmatrix} \tag{3}$$

Combining x_i and $\mathbf{a}_{o,i}$ for batch i into the model input z_i, the model can be assumed generally to be

$$y_i=f(x_i, \mathbf{a}_{o,i})=f(\mathbf{z}_i) \tag{4}$$

where f(•) is the function relating y to x for run i and the relationship between y and x is often determined empirically. The schematic diagram of the operating variables for the processes is shown in Fig. 1.

In the past, studies on modeling of high-mix processes were focused on the linear process model given by

$$y_i=bx_i+\mathbf{a}_{o,i}\mathbf{c} \tag{5}$$

where b refers to gain matrix and $\mathbf{c}$ refers to matrix of tools and products.

There are various methods mentioned in the literature on estimation of $\mathbf{c}$ [7-9]. One method is JADE (just in time adaptive disturbance estimation). Another method is RLS-EFRA (recursive least squares-exponential forgetting and resetting algorithm). Since the process model is linear, the conventional EWMA (exponentially weighted moving average) controller can be used. Even for the linear

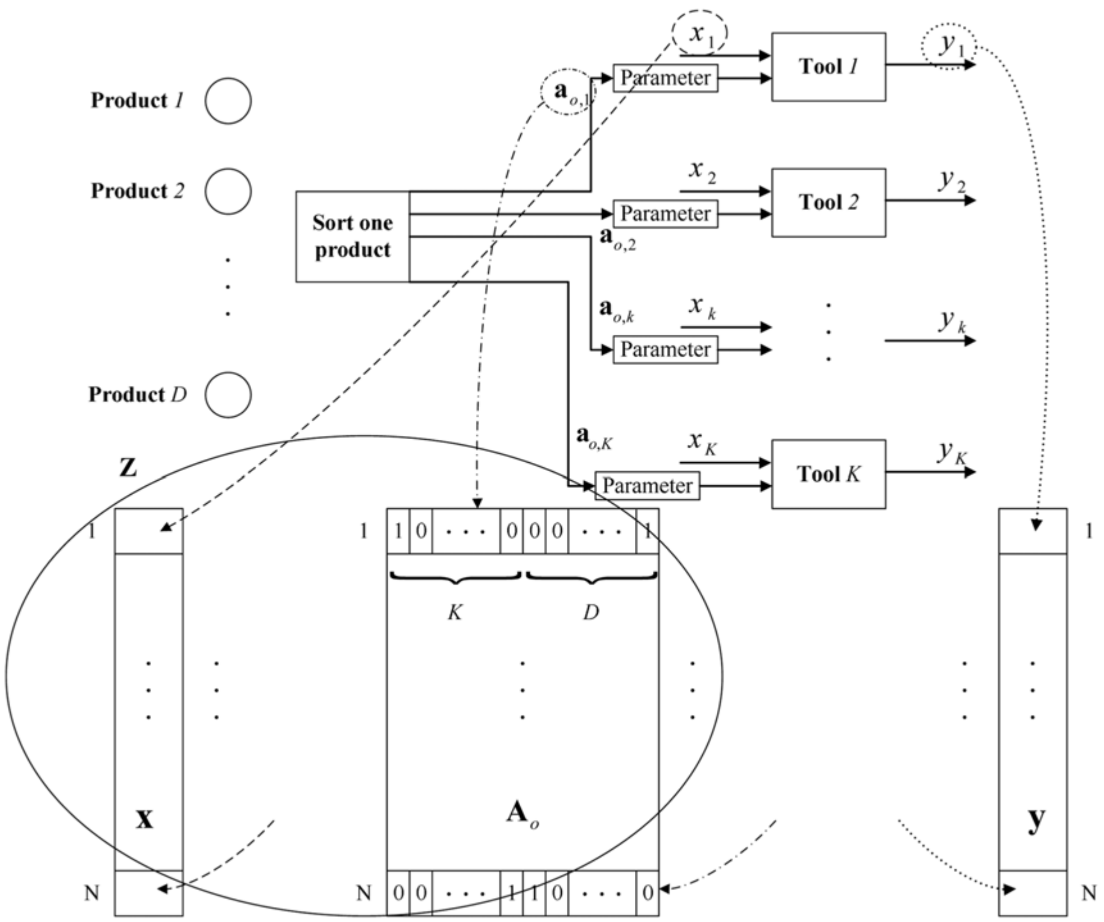

Fig. 1. MTMP variable schematic diagram.

process, the tool and product specific constants are difficult to determine because the context variables are confounded with each other and the system is poorly excited. Furthermore, if the process model is nonlinear, process modeling and control using the EWMA controller coupled with JADE or RLS-EFRA will not be satisfactory.

MODELING OF NONLINEAR MTMP PROCESSES

Designed for nonlinear processes, a novel RtR methodology based on KSVM is developed. KSVM has the excellent generalization performance that the model built using it from limited data would still yield reliable results. KSVM can perform the nonlinear regression since it utilizes a nonlinear mapping of the original variables into a feature space where the linear regression is done. Modeling of the nonlinear system will be discussed first and then followed by the development of the KSVM control algorithm.

Given a training data set of N points $\{\mathbf{z}_i, y_i\}_{i=1}^{N}$ with input data $\mathbf{z}_i \in R^{m_D}$ (m_D=K+D+1) and output data $y_i \in R$, the goal is to estimate a model of the form

$$y(\mathbf{z}) = \sum_{j=1}^{m_p} w_j \varphi_j(\mathbf{z}) + h = \mathbf{w}^T \Phi(\mathbf{z}) + h \tag{6}$$

where m_p is the dimension of $\varphi_j(x)$ space.

The following constrained optimization problem in the primal space is to be solved for the best choice of the weight vector $\mathbf{w} \in R^{m_p}$ and the common bias h.

$$\min_{\omega,b,e} J(\mathbf{w}, \mathbf{e}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + \frac{1}{2}\gamma\sum_{i=1}^{N} e_i^2 \quad \text{s.t.} \quad y_i = \mathbf{w}^T\varphi(\mathbf{z}_i) + h + e_i, \forall i \tag{7}$$

where the error vector $\mathbf{e} \in R^N$, $e_i = y_i - \mathbf{w}^T\varphi(\mathbf{z}_i) - h$, and $\varphi(\square)$: $R^{m_D} \rightarrow R^{m_p}$ is a function which maps the input space into a higher dimensional (possibly infinite dimensional) feature space. γ is a specified trade-off coefficient between a smoother solution and training error weighting and is taken to be 100 in applying SVM in this paper. The constrained optimization problem is reformulated in the dual space by constructing the Lagrangian. The solution of h can be found by solving

$$\begin{bmatrix} 0 & \mathbf{1}_v^T \\ \mathbf{1}_v & \Omega + \gamma^{-1}\mathbf{I} \end{bmatrix} \begin{bmatrix} h \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix} \tag{8}$$

where $\mathbf{y} = [y_1\ y_2\ y_3 \cdots y_N]^T$, $\mathbf{1}_v = [1\ 1\ 1 \cdots 1]^T$, and $\Omega = \varphi(\mathbf{z}_j)^T\varphi(\mathbf{z}_l)$ for j, l=1, …, N. $\alpha = [\alpha_1\ \alpha_2\ \alpha_3 \cdots \alpha_N]^T$ is the vector of the Lagrange multipliers. According to Mercer's theorem, a convenient kernel choice $K(\mathbf{z}_j, \mathbf{z}_l)$ is

$$K(\mathbf{z}_j, \mathbf{z}_l) = \varphi(\mathbf{z}_j)^T\varphi(\mathbf{z}_l), \qquad j, l = 1, \cdots, N \tag{9}$$

The resulting LS-KSVM model for the functional estimation is

$$y(\mathbf{z}) = \sum_{j=1}^{N} \alpha_j K(\mathbf{z}, \mathbf{z}_j) + h \tag{10}$$

where α and h are the least squares solutions to Eq. (8). The least squares SVM with the radial basis kernel function $K(\mathbf{z}_j, \mathbf{z}_l) = \exp(-(\|\mathbf{z}_j - \mathbf{z}_l\|_2^2)/\sigma^2)$ is used in this paper. If LS-SVM is used, the same procedure is followed except for $\Omega = \mathbf{z}_j^T\mathbf{z}_l$ and the kernel function used should be $K(\mathbf{z}_j, \mathbf{z}_l) = \mathbf{z}_j^T\mathbf{z}_l$, k, l=1, ⋯, N. A detailed description of SVM modeling can be found in the book by [10].

DESIGN OF RTR CONTROLLERS IN MTMP

The objective for RtR control is to minimize the square of the predicted error in tool k for batch i+1. The predicted error is the difference between the setpoint and the output for tool k and batch i+1.

$$\min_{x_{k,i+1}} J_k = \min_{x_{k,i+1}} [e_{k,i+1}^T e_{k,i+1}] \tag{11}$$

$e_{k,i+1}$, the error obtained for batch i+1 and tool k is defined to be

$$e_{k,i+1} \equiv y_{k,i+1}^{sp} - \hat{y}_{k,i+1} \tag{12}$$

The KSVM model is modified by adding two constant terms to account for the presence of the unknown disturbances and drifts. These constants are modified each batch in the same manner as the constants in the dEWMA method by

$$y_{k,i+1} = y_{k,i} + a_{k,i+1} + d_{k,i+1} = \sum_{n=1}^{N} \alpha_j \exp\left[\frac{-\|\mathbf{z}_n - \mathbf{z}_{k,i+1}\|^2}{\sigma^2}\right] + h + a_{k,i+1} + d_{k,i+1} \tag{13}$$

The term $a_{k,i+1}$ refers to bias correction and term $d_{k,i+1}$ refers to drift correction. Both terms $a_{k,i+1}$ and $d_{k,i+1}$ are corrected at the end of each batch for tool k using the form below.

$$\begin{aligned} a_{k,i+1} &= \lambda_1(y_{k,i} - \hat{y}_{k,i}) + (1-\lambda_1)a_{k,i} & 0 \le \lambda_1 \le 1 \\ d_{k,i+1} &= \lambda_2(y_{k,i} - \hat{y}_{k,i} - a_{k,i}) + (1-\lambda_2)d_{k,i} & 0 \le \lambda_2 \le 1 \end{aligned} \tag{14}$$

The control objective, minimization of square of the predicted error $e_{k,i+1}$ in tool k for batch i+1, becomes

$$\min_{z_{k,i+1}} \left(y_{k,i+1}^{sp} - \sum_{n=1}^{N} \alpha_n \exp\left(\frac{-\|\mathbf{z}_n - \mathbf{z}_{k,i+1}\|^2}{\sigma^2}\right) - h - a_{k,i+1} - d_{k,i+1} \right)^2 \tag{15}$$

Taking the derivative of the bracketed expression in Eq. (15) to zero, we get

$$\frac{\partial J}{\partial x_{k,i+1}} = -2\frac{\partial(\hat{y}_{k,i+1})}{\partial x_{k,i+1}}(y_{k,i+1}^{sp} - \hat{y}_{k,i+1} - a_{k,i+1} - d_{k,i+1}) = 0 \tag{16}$$

It is not possible to obtain an analytical expression for $\mathbf{z}_{k,i+1}$ based on the above equation. With the first order Taylor approximation, an estimate of the optimum $\mathbf{z}_{k,i+1}$ satisfying Eq. (15) can be obtained numerically. Using Eqs. (17)-(19), an estimate of the optimum control input $x_{k,i+1}$ is obtained and this input is augmented with the context variables $\mathbf{a}_{o,k}$ to get estimate of the optimum $\mathbf{z}_{k,i+1}$.

$$x_{k,i+1}\big|_{est} = x_{k,i} - \frac{\kappa\sigma^2}{2\sum_{j=1}^{N}\alpha_j\beta_j(x_j - x_{k,i})} \tag{17}$$

where

$$\beta_j = \exp\left(\frac{-(x_j - x_{k,i}) - \|\mathbf{a}_{0,j}^t - \mathbf{a}_{k,i}^t\|^2 - \|\mathbf{a}_{0,j}^p - \mathbf{a}_{k,i}^p\|^2}{\sigma^2}\right) \tag{18}$$

and

$$\kappa = y_{k,i+1}^{sp} - \sum_{j=1}^{N}\alpha_j\beta_j - h - a_{k,i+1} - d_{k,i+1} \tag{19}$$

Eq. (17) provides an estimate of the control input needed. In this paper, the controller implementing the action is called linear approximate KSVM controller. This estimate together with $\mathbf{a}_{o,k}$ can serve

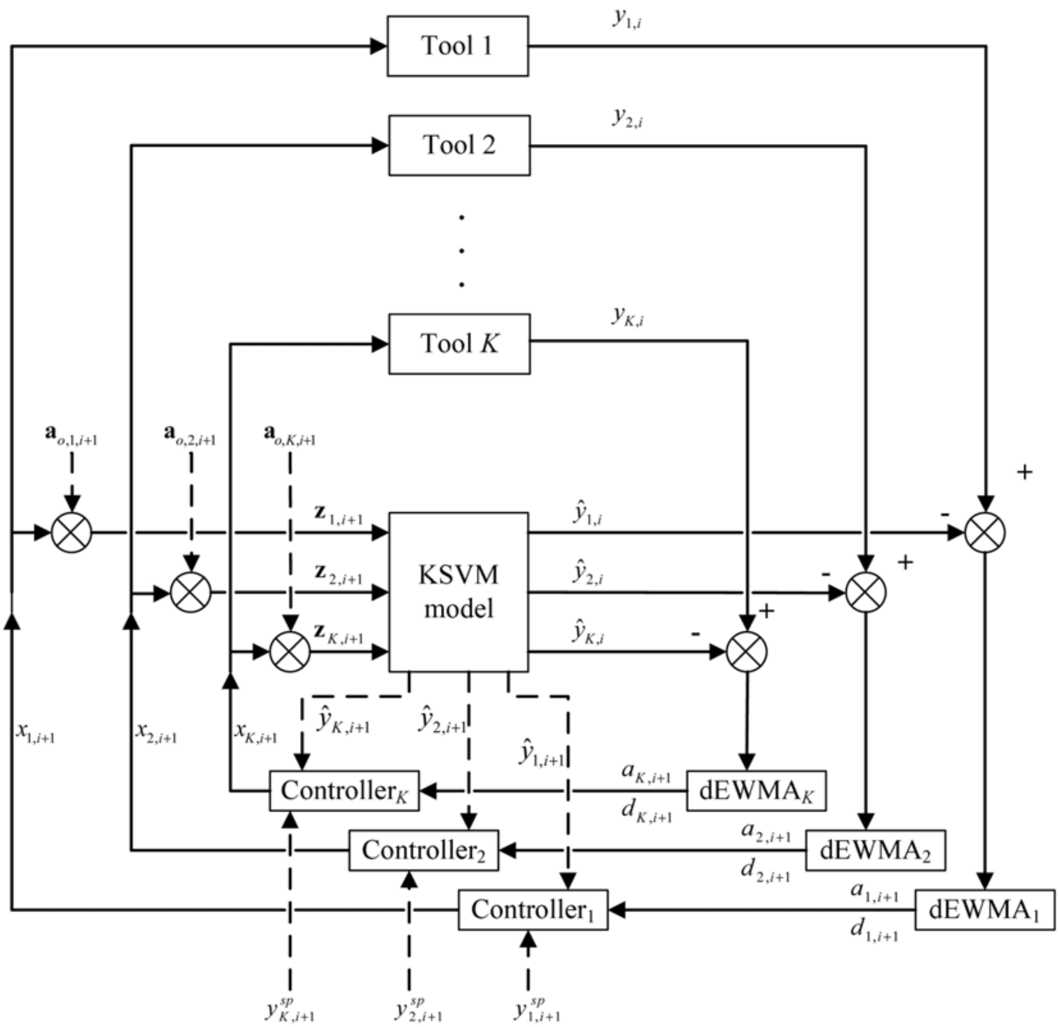

Fig. 2. Schematic diagram of MTMP control loops.

as an initial estimate of $\mathbf{z}_{k,i+1}$ and the objective function presented in Eq. (15) is optimized to determine the optimum $\mathbf{z}_{k,i+1}$. The controller implementing the action is called KSVM controller. A schematic drawing of the control system is given in Fig. 2.

ILLUSTRATIVE EXAMPLES

The shallow trench isolation (STI) process is used here. The key steps of STI processes involve etching of a pattern of trenches in the substrate material, depositing of silicon or other dielectric materials to fill the trenches and removing excess dielectric using chemical-mechanical planarization [11]. According to the graph shown by Kazuhiro, the etching depth y_i can be a nonlinear function of the etching time x_i for batch i [12]. The nonlinear relationship of the STI etching process is assumed as follows.

$$y_i=f(x_i, \mathbf{a}_{o,i})=-1.8287x_i^2+10.283x_i+\mathbf{a}_{o,i}\mathbf{c}+T_{SN,d}+g_{i,k}+\varepsilon_{i,k} \tag{20}$$

where y_i is the output for *i*th batch and x_i is the control input for the *i*th batch. Let $\mathbf{a}_{o,i}$ represent the context variable contribution for the *i*th batch. $g_{i,k}$ is an optional term accounting for the presence of drift due to tool deterioration for tool k at batch i. It is assumed that for batch i, tool k is used to produce product d. In testing for the effect of drift due to tool deterioration, the drift term expression $g_{i,k}$ in terms of batch number i for any tool k is given to be

$$g_{i,k}=\frac{i-20}{19} \tag{21}$$

$T_{SN,d}$ is an initial depth of product d and it is a random constant drawn from Gaussian distribution of mean −0.0117 and standard deviation of 1.02. In this example, three tools are used to produce four products, and bias matrix **c** contains both the tool and product biases and is given to be

$$\mathbf{c}=[5\ 6\ 7\ 15\ 25\ 35\ 45]^T \tag{22}$$

The first three constants refer to the tool bias constants of tool 1, 2 and 3 and the last four constants refer to the product bias constants of product 1, 2, 3 and 4. White noise $\varepsilon_{i,k}$ at batch i and tool k is added to y_i using Gaussian distribution with zero mean and the standard deviation of 1.0, 1.2 and 1.1 are given for tool 1, 2 and 3, respectively.

Data has been generated for 50 batches. The product produced is changed every 10 batches. The production schedule is shown in Table 1. The first 30 batches of the data generated are used as a training data set and the last 20 batches are used as a test data set. Based

Table 1. Production schedule for model building

Tool\ Period	Period I	Period II	Period III	Period IV	Period V
Tool 1	Product 1	Product 1	Product 3	Product 4	Product 3
Tool 2	Product 3	Product 4	Product 1	Product 4	Product 2
Tool 3	Product 1	Product 1	Product 2	Product 2	Product 1

Table 2. Production schedule for control purposes

Tool\ Period	Period I	Period II	Period III	Period IV	Period V
Tool 1	Product 1	Product 2	Product 3	Product 4	Product 1
Tool 2	Product 2	Product 3	Product 1	Product 1	Product 2
Tool 3	Product 3	Product 4	Product 1	Product 2	Product 3

on the training data, the KSVM method predicts correctly the process outputs given the control inputs and the context variables. The performance of the control system is evaluated in the presence of persistent drift. For the testing case, a new, specified production schedule is presented in Table 2. The setpoints for producing the different products are 23, 30, 42 and 50 for products 1, 2, 3 and 4, respectively. The effectiveness of the following control designs is compared with the optimal KSVM controller:

1. Linear approximate KSVM controller: to test the effectiveness of using linear approximate expression instead of the optimum control input
2. EWMA controller coupled with JADE or RLS-EFRA and SVM: to test the effectiveness of process nonlinearity on linear process controllers
3. Context-free KSVM controller: to test the effectiveness of the context variables

1. Case 1: Comparison of Performance between Linear Approximate KSVM and KSVM RtRs

For the case, the linear estimate is used to design the optimal control action. It is of interest to determine whether more computations

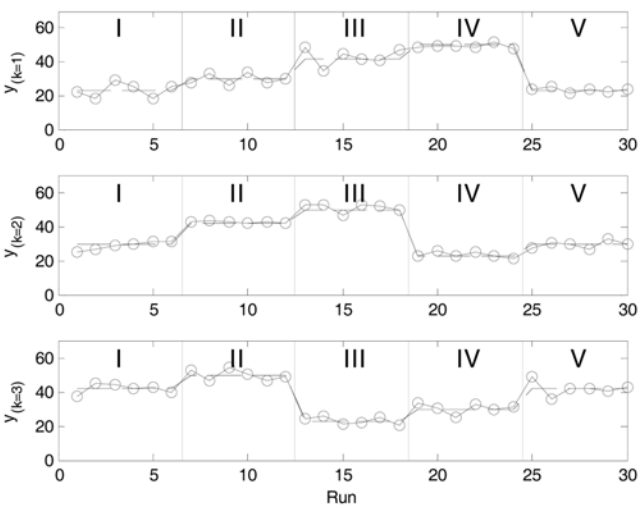

Fig. 3. Responses of linear approximate KSVM controller to drifts in the STI process: —○: control data, ——: set point.

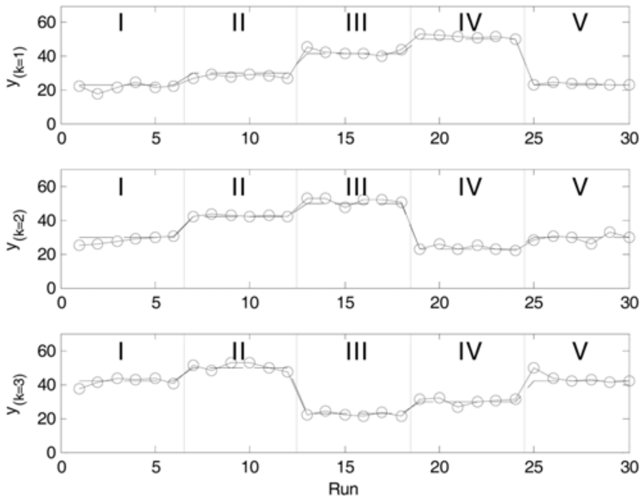

Fig. 4. Responses of KSVM controller to drifts in the STI process: —○: control data, ——: set point.

are warranted to obtain the optimal control action as presented in Eq. (15) or the use of a linear estimate of the control input in Eq. (17) will be sufficient for control purposes. The response of linear approximate and KSVM controller to drifts due to aging is plotted in Figs. 3 and 4 for linear approximate KSVM and KSVM controllers. KSVM tracks the setpoints slightly closer than linear approximate KSVM. The dEWMA constants are used for the three tools:

tool 1	λ_1=0.5	λ_2=0.05
tool 2	λ_1=0.5	λ_2=0.1
tool 3	λ_1=0.35	λ_2=0.1

The performance of KSVM is slightly better than that of linear approximate KSVM controller since the former is able to track the setpoint closer and faster.

2. Case 2: Comparison of Performance between JADE, RLS-EFRA and KSVM Control Designs

A comparison of the performance is made between EWMA and KSVM based control designs. It determines whether the control action of EWMA coupled with JADE and RLS-EFRA controllers would be sufficient to control the process so that the process outputs track and eventually match the setpoint within the appropriate production period in the presence of unknown disturbances and drifts. JADE and RLS-EFRA would yield linear models. The response of EWMA and KSVM control to drifts due to aging is plotted in Figs. 5(a), 5(b) and Fig. 4, respectively. Offsets are observed for EWMA/JADE controller and EWMA/RLS-EFRA controllers as shown in Figs. 5(a) and 5(b), specifically: Tool 1 period I, Tool 2 Period II and Tool 3 Period I and II. This comparison underscores the importance of having a more accurate process model.

3. Case 3: Comparison of Performance between KSVM with Context and Context-free Based Control Designs

The comparison of the performance is made between threaded model-based control and non-threaded model-based control and their responses to the unknown disturbances and drifts. The responses of context-free KSVM control to drifts due to aging are plotted in Fig. 6. It is sluggish and the offsets prevail over most of the production period. It indicates that knowledge of context variables is critical

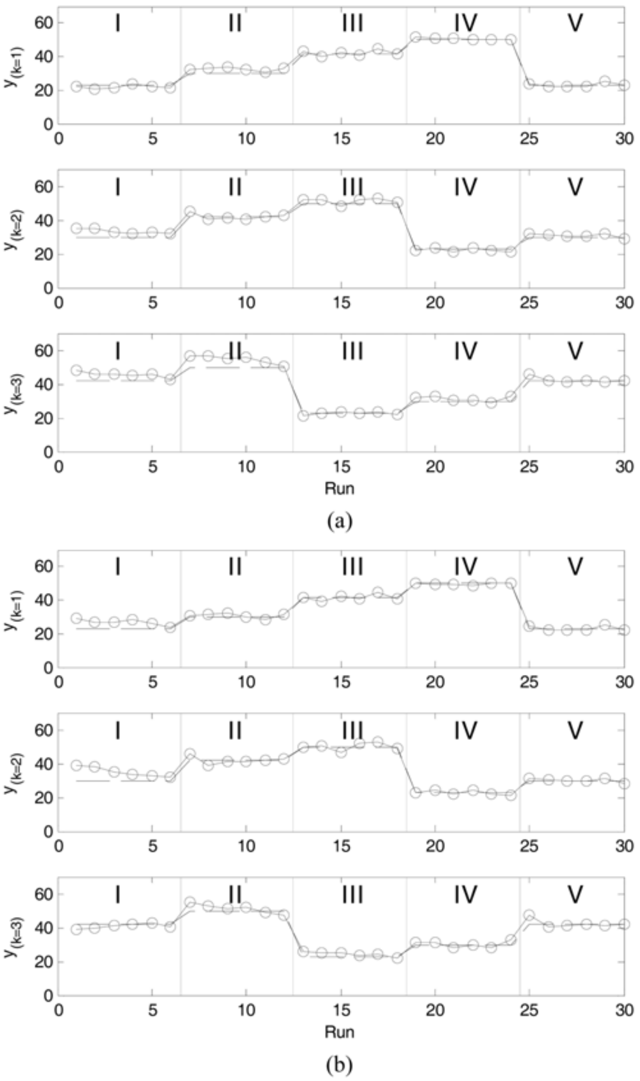

Fig. 5. Responses of linear controllers to drifts in the STI process: (a) JADE (b) RLS.

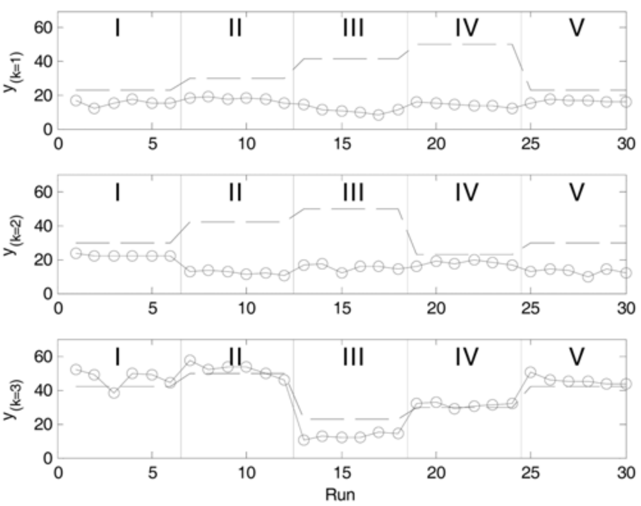

Fig. 6. Responses of context free KSVM control to drifts in the STI process: —○: control data, ——: set point.

to the success of MTMP controllers.

CONCLUSIONS

A new RtR control methodology that is applicable for nonlinear semiconductor processes is proposed. The nonlinear modeling is done using KSVM which utilizes the radial basis function as the kernel function. To eliminate the effects of the unknown disturbances and drifts, the control law for the KSVM controller is derived. The effectiveness of KSVM as a modeling technique and its controller is demonstrated through an STI simulation study. The performance of the KSVM control algorithm is superior to the other MTMP control algorithms in controlling processes. The effect of using other kernel functions on the performance of KSVM is a topic of our future research. Since process nonlinearity comes in many forms, it would be helpful to establish some criteria to guide the selection of the form of kernel functions for nonlinear modeling.

ACKNOWLEDGEMENTS

The authors wish to express their sincere gratitude to the Center-of-Excellence (COE) Program on Membrane Technology from the Ministry of Education (MOE), R.O.C., and to the project Toward Sustainable Green Technology in the Chung Yuan Christian University, Taiwan, under grant CYCU-98-CR-CE for their financial support.

REFERENCES

1. S. J. Qin, G. Cherry, R. Good, J. Wang and C. A. Harrison, *J. Process Control*, **16**, 179 (2006).
2. E. D. Castillo and A. M. Hurwitz, *J. Quality Technol.*, **29**, 184 (1997).
3. A. J. Pasadyn and T. F. Edgar, *IEEE Transactions on Semiconductor Manufacturing*, **18**, 592 (2005).
4. V. Vapnik, *Estimation of Dependences Based on Empirical Data* [in Russian] 1979. Nauka, Moscow (English translation: Springer-Verlag, New York, 1982).
5. V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York (1995).
6. J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor and J. Vandewalle, *Least Squares Support Vector Machines*, Singapore, World Scientific (2002).
7. C. A. Bode, J. Wang, Q. P. He and T. F. Edgar, *Annual Reviews in Control*, **31**, 241 (2007).
8. J. Wang, Q. P. He and T. F. Edgar, *American Control Conference*, 3636 (2007).
9. J. Wang, Q. P. He and T. F. Edgar, *J. Process Control*, **19**, 443 (2009).
10. J. A. K. Suykens, J. D. Brabanter, L. Lukas and J. Vandewalle, *Neurocomputing*, **48**, 85 (2002).
11. S. Gaddam and M. W. Braun, *Advanced Semiconductor Manufacturing Conference and Workshop*, 17 (2005).
12. K. Miwa, T. Inokuchi, T. Takahashi, A. Oikawa and K. Imaoka, *IEEE Transactions on Semiconductor Manufacturing*, **18**, 517 (2005).