

A Virtual Metrology Model Based on Recursive Canonical Variate Analysis with Applications to Sputtering Process

Tian-hong Pan^a, David Shan-Hill Wong^b, Shi-Shang Jang^b

^aSchool of Electrical and Information Engineering, Jiangsu University, Zhenjiang, Jiangsu, 212013, China

^bChemical Engineering Department, National Tsing-hua University, Hsin-Chu, 30047, Taiwan, E-mail: ssjang@mx.nthu.edu.tw

In data driven process monitoring, soft-sensor, or virtual metrology (VM) model is often employed to predict product's quality variables using sensor variables of the manufacturing process. Partial least squares (PLS) is commonly used to achieve this purpose. However PLS seeks the direction of maximum co-variation between process variables and quality variables. Hence, a PLS model may include the directions represent variations in the process sensor variables that are irrelevant to predicting quality variables. In the case, when direction of sensor variables' variations that are most influential to quality variables is near orthogonal to direction of largest process variations, a PLS model will lack generalization capability. In contrast to PLS, canonical variate analysis (CVA) identifies a set of basis vector pairs which would maximize the correlation between input and output. Thus, it may uncover complex relationships that reflect the structure between quality variables and process sensor variables. In this work, an adaptive VM based on recursive CVA (RCVA) is proposed. Case study on an industrial sputtering process demonstrates the capability of CVA-based VM model compared to PLS-based VM model.

1. Introduction

Modern industrial processes, continuous or batch, are usually equipped with a large number of sensors that provide process variables data such as pressure, temperature, spectroscopic signals, heat or power supplied, etc. Such process variables data can be used for process monitoring and fault diagnosis using multivariate statistical analysis methods such as principle component regression (PCR), partial least squares (PLS) and canonical variate analysis (CVA). However, such process variables data are not direct indicator of final product quality. Examples of quality variables in continuous chemical process include molecular weight and distribution of polymers, purity of distillates, etc. In multi-step batch-based processes, such as semiconductor or thin-film-transistor liquid-crystal-display (TFT-LCD) manufacturing, intermediate product qualities include product state variables such as film thickness, critical dimension etc., and final quality indicators include electrical characteristics such as sheet resistance, threshold voltage, etc. (Mays and Spanos 2003). Since product quality measurements may be time-consuming and expensive, they can only be sampled and provided in a less frequent, time-delayed manner. For both monitoring and controlled purposes, it is therefore

desirable that such quality data can be predicted using process variables data. Such an approach is well known as soft-sensor or VM. Examples of VM applications can be found in both continuous chemical processing industry (Kano and Nakagawa, 2008) as well as multi-step batch-wise assembly line processes (Zeng and Spanos, 2009).

One approach to development of VM models is to use non-parametric models such as neural networks (Radhakrishnan and Mohamed, 2004). Neural network models are capable of representing complex non-linear functions, but they usually lack generalization ability unless special attentions were paid to variable selection, architecture determination and data screening and compression (Lin et al, 2009). Alternatively, linear model, such as PLS is also commonly used to construct VM models (Sharmin, 2006). Such models are capable of overcoming the problem of high dimensionality and collinearity in the process variables data. However PLS seeks the direction of maximum co-variation between process variables and quality variables. Hence, a PLS model may include the directions represent variations in the process sensor variables that are irrelevant to predicting quality variables. In the case, when direction of sensor variables' variations that are most influential to quality variables is near orthogonal to direction of largest process variations, a PLS model will lack generalization capability. In contrast to PLS, CVA identifies a set of basis vector pairs which would maximize the correlation between input and output. Thus, it may uncover complex relationships that reflect the structure between quality variables and process sensor variables. In this work, an adaptive VM based on RCVA is proposed. Case study on an industrial example demonstrates the capability of CVA-based VM model.

2. Methods

2.1 Standard CVA

Let us define a normalized (zero mean and unit variance) set of observed input process variable data $\mathbf{X} = [\mathbf{x}(1), \mathbf{x}(2) \cdots, \mathbf{x}(K)]^T$, and the corresponding normalized observed output quality data $\mathbf{Y} = [\mathbf{y}(1), \mathbf{y}(2) \cdots, \mathbf{y}(K)]^T$, where $k = 1 \cdots K$ is the sampling index.

Assume that there are $p = 1 \cdots P$ input process variables: $\mathbf{x}(k) = [x_1(k) \cdots, x_p(k)]$, and that there are $q = 1 \cdots Q$ output quality variables: $\mathbf{y}(k) = [y_1(k), \cdots, y_q(k)]$.

CVA aims to find a set of canonical vectors $\mathbf{W}_x = [\mathbf{w}_x^1, \cdots, \mathbf{w}_x^N] \in \mathfrak{R}^{P \times N}$ and $\mathbf{W}_y = [\mathbf{w}_y^1, \cdots, \mathbf{w}_y^N] \in \mathfrak{R}^{Q \times N}$ that maximizes the correlation between the output quality variables and input process variables that are orthogonal to each other.

$$\begin{aligned} \max \left\{ \rho^i = \frac{(\mathbf{w}_x^i)^T \boldsymbol{\Sigma}_{xy} (\mathbf{w}_y^i)}{\sqrt{(\mathbf{w}_x^i)^T \boldsymbol{\Sigma}_{xx} (\mathbf{w}_x^i)} \sqrt{(\mathbf{w}_y^i)^T \boldsymbol{\Sigma}_{yy} (\mathbf{w}_y^i)}} \right\} \\ \text{s.t. } (\mathbf{w}_x^i)^T \boldsymbol{\Sigma}_{xx} (\mathbf{w}_x^i) = 1 = (\mathbf{w}_y^i)^T \boldsymbol{\Sigma}_{yy} (\mathbf{w}_y^i) \\ (\mathbf{w}_x^i)^T \boldsymbol{\Sigma}_{xx} (\mathbf{w}_x^j) = 0 = (\mathbf{w}_y^i)^T \boldsymbol{\Sigma}_{yy} (\mathbf{w}_y^j) \quad j \neq i, i, j = 1, 2, \cdots, N \end{aligned} \quad (1)$$

This can be done by solving the eigenvalue problem:

$$\begin{bmatrix} \Sigma_{xx} & \mathbf{0} \\ \mathbf{0} & \Sigma_{yy} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} & \Sigma_{xy} \\ \Sigma_{yx} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{w}_x \\ \mathbf{w}_y \end{bmatrix} = \rho \begin{bmatrix} \mathbf{w}_x \\ \mathbf{w}_y \end{bmatrix} \quad (2)$$

where ρ is canonical correlation coefficients, i, j is the index of canonical vectors, Σ_{xx} , Σ_{yy} are the covariance matrices of input and output, respectively. $\Sigma_{yx} = (\Sigma_{xy})^T$ is the cross covariance matrix between input and output.

2.2 VM model based on CVA

After finding the canonical vectors, a model prediction of the quality variables at the time point k^{th} can be written as:

$$\begin{aligned} \tilde{\mathbf{y}}(k) &= (\mathbf{W}_y)^{-1} (\mathbf{B} \bullet \mathbf{x}(k) \mathbf{W}_x) \\ \mathbf{B} &= \text{diag}(b^1, b^2, \dots, b^N) \end{aligned} \quad (3)$$

The coefficients of this model b^1, b^2, \dots, b^N can be obtained by standard least square estimation

$$b^i = (\mathbf{X}^T \mathbf{w}_x^i)^T \mathbf{Y}^T \mathbf{w}_y^i \text{pinv} \left[(\mathbf{X}^T \mathbf{w}_x^i)^T (\mathbf{X}^T \mathbf{w}_x^i) \right] \quad (4)$$

2.3 Recursive form of the CVA based VM model

In on-line applications, it is desirable that the VM model or monitoring model will be implemented in a recursive manner with forgetting factors, but without the need of recalling past training data (Li et al., 2000). Using equations (2) to (4), we can see that estimation of the loading factors and the regression coefficients can be recursively estimated by the procedure of normalizing data and calculation of covariance matrix can be carried out recursively without recalling past training data.

Let us denote that observed, i.e. un-normalized, response and regressor variables at time point k^{th} by $\hat{\mathbf{y}}(k)$ and $\hat{\mathbf{x}}(k)$ respectively. A recursive form of the mean vectors and standard deviation vectors are given as following (Lee and Lee, 2008):

$$\boldsymbol{\mu}_x(k) = \frac{1 - \alpha_x}{1 - \alpha_x^k} \hat{\mathbf{x}}(k) + \alpha_x \frac{1 - \alpha_x^{k-1}}{1 - \alpha_x^k} \boldsymbol{\mu}_x(k-1) \quad (5)$$

$$\boldsymbol{\sigma}_x^2(k) = \frac{(1 - \beta_x)}{(1 - \beta_x^k)} \left[\text{diag}(\hat{\mathbf{x}}(k) - \boldsymbol{\mu}_x(k)) (\hat{\mathbf{x}}(k) - \boldsymbol{\mu}_x(k))^T \right] + \beta_x \frac{(1 - \beta_x^{k-1})}{(1 - \beta_x^k)} \boldsymbol{\sigma}_x^2(k-1) \quad (6)$$

where α_x and β_x are the forgetting factors, $\boldsymbol{\mu}_x(k)$ and $\boldsymbol{\sigma}_x(k)$ are the mean vectors and standard deviation vectors of $\hat{\mathbf{x}}(k)$. When the k becomes large, the above equations can be simplified to the following forms

$$\boldsymbol{\mu}_x(k) \approx (1 - \alpha_x) \hat{\mathbf{x}}(k) + \alpha_x \boldsymbol{\mu}_x(k-1) \quad (7)$$

$$\boldsymbol{\sigma}_x^2(k) \approx (1 - \beta_x) \left[\text{diag}(\hat{\mathbf{x}}(k) - \boldsymbol{\mu}_x(k)) (\hat{\mathbf{x}}(k) - \boldsymbol{\mu}_x(k))^T \right] + \beta_x \boldsymbol{\sigma}_x^2(k-1) \quad (8)$$

Similarly, the mean $\boldsymbol{\mu}_y(k)$ and variance $\boldsymbol{\sigma}_y^2(k)$ vectors can also be deduced as the recursive forms:

$$\boldsymbol{\mu}_y(k) \approx (1 - \alpha_y) \hat{\mathbf{y}}(k) + \alpha_y \boldsymbol{\mu}_y(k-1) \quad (9)$$

$$\boldsymbol{\sigma}_y^2(k) \approx (1 - \beta_y) \left[\text{diag}(\hat{\mathbf{y}}(k) - \boldsymbol{\mu}_y(k)) (\hat{\mathbf{y}}(k) - \boldsymbol{\mu}_y(k))^T \right] + \beta_y \boldsymbol{\sigma}_y^2(k-1) \quad (10)$$

where α_y and β_y are the forgetting factors.

The normalized response and regressor variables thus can be computed as:

$$\mathbf{x}(k) = (\boldsymbol{\sigma}_x)^{-1} [\hat{\mathbf{x}}(k) - \boldsymbol{\mu}_x(k)] \quad (11)$$

$$\mathbf{y}(k) = (\boldsymbol{\sigma}_y)^{-1} [\hat{\mathbf{y}}(k) - \boldsymbol{\mu}_y(k)] \quad (12)$$

Similarly, the covariance matrices can also be recursively calculated as:

$$\boldsymbol{\Sigma}_{xx}(k) = (1 - \beta_x) \mathbf{x}(k) (\mathbf{x}(k))^T + \beta_x \boldsymbol{\Sigma}_{xx}(k-1) \quad (13)$$

$$\boldsymbol{\Sigma}_{yy}(k) = (1 - \beta_y) \mathbf{y}(k) (\mathbf{y}(k))^T + \beta_y \boldsymbol{\Sigma}_{yy}(k-1) \quad (14)$$

$$\boldsymbol{\Sigma}_{xy}(k) = (1 - \beta_{xy}) \mathbf{x}(k) (\mathbf{y}(k))^T + \beta_{xy} \boldsymbol{\Sigma}_{xy}(k-1) \quad (15)$$

where β_{xy} is the forgetting factors,

Given these covariance vectors, the canonical correlation loading vectors $\mathbf{W}_x(k)$ and $\mathbf{W}_y(k)$ can be obtained by the optimization procedure described in equation (2) and the correlation coefficients ρ^i and the model coefficients b^i can be obtained by equations (1) and (4).

3. Application to a Sputtering Process

3.1 Process and data description

In this section, VM models were developed for a sputtering process in a local TFT-LCD manufacturing facility. In this process, ions in the plasma are accelerated towards a target. Atoms of the target are sputtered and ejected into the plasma and eventually deposited on the surface of a glass substrate (Pai et al, 2009). There are two sputtering steps in which different atoms are sputtered and deposited. The process operation is monitored by six kinds of equipment sensors which recorded the supplied power, voltage, pressure and composition of the residual gases etc. For some of these sensing variables there may be several probes at different locations of the chamber. Each of these probes reported data at intervals of micro-seconds. These profile data are summarized according to operating characteristics of each step. Eventually we are supplied with 21 characteristics process variables, i.e. $\hat{\mathbf{x}}(k) = [\hat{x}_1(k), \hat{x}_2(k), \dots, \hat{x}_{21}(k)]$. The quality variable of concern is the sheet resistance of the glass substrates at 9 different locations $\hat{\mathbf{y}}(k) = [\hat{y}_1(k), \hat{y}_2(k), \dots, \hat{y}_9(k)]$. 110 samples covering four months were collected for VM modeling after data preprocessing. Various events

including preventive maintenance, idle operation and change of recipes have occurred during this period.

3.2 Results

The first 21 samples were used as the initial batch for training and the rest are predicted by RCVA model. The forgetting factors were set as: $\alpha_x = \alpha_y = 0.99$, $\beta_x = \beta_{xy} = \beta_y = 0.9$. Nine CVA components are selected.

Table 1 demonstrates the accuracies of recursive tracking of the values of y using RCVA models. It can be seen that the accuracy of RCVA model is much good enough. The R^2 of the RCVA predictions varied from 48 to 65% which is consistent with the modeling error achieved.

Table 1: The prediction accuracy of RCVA VM models

position		1	2	3	4	5	6	7	8	9
RCVA	RMSE (10^{-2})	1.54	1.51	1.49	1.56	1.61	1.86	1.48	2.03	1.25
	MAPE (%)	1.08	1.14	1.13	1.18	1.20	1.36	1.15	1.44	0.95
	R^2 (%)	54.68	58.69	62.53	57.01	64.63	54.50	59.16	48.15	62.40

p.s.

$$RMSE_q = \sqrt{\frac{1}{K} \sum_{k=1}^K (\hat{y}_q(k) - \tilde{y}_q(k))^2}, MAPE_q = \frac{1}{K} \sum_{k=1}^K \left| 1 - \frac{\tilde{y}_q(k)}{\hat{y}_q(k)} \right| \times 100\%$$

$$\hat{\mu}_q = \frac{1}{K} \sum_{k=1}^K \hat{y}_q(k), SST_q = \sum_{k=1}^K (\hat{y}_q(k) - \hat{\mu}_q)^2, SSE_q = \sum_{k=1}^K (\hat{y}_q(k) - \tilde{y}_q(k))^2,$$

$$R^2_q = \left(1 - \frac{SSE_q}{SST_q} \right) \times 100\%$$

$$q = 1, 2, \dots, 9$$

However, the predicted accuracy of RCVA and RPLS are also influenced by the forgetting factor. As mentioned by Sang et al. (Sang et al., 2006), the optimal value of the forgetting factor which varies significantly depending on the rate of process change can improve the performance of adaptive algorithm. In this work, we only compare the performance of RPLS and RCVA under the same forgetting factors.

4. Conclusions

In this work, we have presented a detailed approach for constructing a VM model using a recursive CVA. We have demonstrated the advantages of this approach using an application to a sputtering process. The advantages are due to the fact that CVA

captured the directions of maximum correlation between process sensor variable input x and quality variable y of the VM model. A PLS model will include variations in x that are large but irrelevant to quality predictions. This is especially important when VM models are developed for processes in which process sensors are comprehensively placed irrespective of their relevance of final quality. In such applications, the generalization ability of a CVA-based VM model is superior.

Acknowledgment

The authors would like to thank the financial support provided by National Science Council under the grant NSC 098-2811-E-007-051 and National Nature Science Foundation of China under Grant 60904053.

References

- Kano M., Nakagawa Y., 2008, Data-based process monitoring, process control, and quality improvement: Recent developments and applications in steel industry, *Computers and Chemical Engineering*, 32, 12-24.
- Lee C., Lee I.B., 2008, Adaptive monitoring statistics with state space model updating based on canonical variate analysis, *Korean Journal of Chemical Engineering*, 25, 203-208.
- Li W.H., Yue H.H., Valle-Cervantes S., Qin S.J., 2000, Recursive PCA for adaptive process monitoring, *Journal of Process Control*, 10, 471-486.
- Lin T.H., Cheng F.T., Wu W.M., Kao C.A., Ye A.J., Chang F.C., 2009, NN-based key-variable selection method for enhancing virtual metrology accuracy, *IEEE Transactions on Semiconductor Manufacturing*, 22, 204-211.
- Mays G.S., Spanos C.J., 2003, *Fundamentals of semiconductor manufacturing and process control*, Hoboken New Jersey: Wiley.
- Pai P.F., Wu T.M., Lin K.P., Yang S.L., 2009, Diagnosing faulty products in TFT-LCD manufacturing processes by support vector machines with principal components analysis, In *Proc. 2009 WRI Global Congress on Intelligent Systems*, pp. 413-417.
- Radhakrishnan V.R., Mohamed A.R., 2004, Neural networks for the identification and control of blast furnace hot metal quality, *Journal of Process Control*, 10, 509-524.
- Sang W.C., Elaine B.M., Morris A.J., Lee I.B., 2006, Adaptive multivariate statistical process control for monitoring time-varying processes, *Ind. Eng. Chem. Res.*, 45, 3108-3118.
- Sharmin R., Sundararaj U., Shah S., Griend L.V., Sun Y.J., 2006, Inferential sensors for estimation of polymer quality parameters: Industrial application of a PLS-based soft sensor for a LDPE plant, *Chemical Engineering Science*, 61, 6372-6384.
- Zeng D., Spanos C.J., 2009, Virtual metrology modeling for plasma etch operations, *IEEE Transactions on Semiconductor Manufacturing*, 22, 419-431.