# Online Predictive Maintenance Approach for Semiconductor Equipment

Ming Luo[1], Zhao Xu[2], Hian Leng Chan[1], Marjan Alavi[2]

[1]*Singapore Institute of Manufacturing Technology, Singapore*

*mluo@simtech.a-star.edu.sg*

[2]*Nanyang Technology University, Singapore*

*Abstract*— **In this paper, an online predictive maintenance approach is proposed for monitoring health of semiconductor equipment. It includes two phases, the first is online prediction of the health indicator and the second phase is the classification of the indicator to one of the health states for making maintenance decisions. Kernel recursive least square (KRLS) algorithm is used for online prediction which is computational efficient. The health states of the equipment can be defined based on the requirement specification for the equipment maintenance. The classification is used in the second stage based on the prediction results come from the first stage. The approach is tested with a simulated dataset from a semiconductor tool and results show a relative high accuracy can be achieved with a satisfactory computational efficiency.**

***Keywords***— **fault prediction, kernel, recursive least square, semiconductor equipment, predictive maintenance**

## I. INTRODUCTION

Maintenance management is an essential task in production to keep equipment in good condition, reduce high operational cost caused by unexpected breakdown [1] in order to remain competitive and profitable. The analysis of system faults and prediction of system failure are becoming more and more important in the manufacturing.

Currently, majority semiconductor companies conduct frequent calibration and preventive maintenance (PM) to ensure high equipment availability and reliability. However, a constantly changing landscape of devices to corresponding to customer demands requires equipment work in the condition of high mixed recipes in production line. This challenges the traditional fixed time based PM and increase the rate of corrective maintenance (CM) which are more costly. CM repair machine when the equipment becomes incapable of further operation. The occurrence of failure not only affects production schedule but product quality. Preventive Maintenance (PM) undergoes maintenance scheduling based on a fixed time from expert's experience. An efficient and effective approach to monitor the health state of equipment and predict impending failure has long been an interest of researcher communities as well as industries. If information is available for predicting when the equipment failure is to occur, a production manager is able to schedule maintenance and production proactively hence reduce the cost of damages due to corrective activities. Many failure prediction techniques have been investigated to assist maintenance decision making. The most common way to detect early sign of failure is to use additional sensors on crucial parts of the equipment. If the mechanism of failure is well known and understood to the maintenance engineer, this technique is an effective method. However, the cost of this method could be high. Moreover, in most industrial cases, it is not possible to add additional sensors. Machine learning based on available information for equipment condition assessment is variable solution when detection from sensor is not feasible.

Recently, kernel-based algorithms have been a topic of considerable interest in the machine learning community for online prediction of time-series [2-6]. These kernel-learning algorithms bridge closely two important areas of adaptive filtering and neural networks, and they embody two important methodologies of error-correction learning and memory-based learning clearly. There are also some kernel-based fault detection and fault diagnose methods which are successfully applied to industrial equipments [7-11]. Kernel recursive least squares (KRLS) algorithm [2] is an efficient online learning algorithm which has already used in some applications. In [12], KRLS is applied to online anomaly detection which raises an alarm immediately upon detecting various kinds of network anomalies in High-speed backbones. KRLS algorithm is also applied to a nonlinear channel identification problem, which typically appears in satellite communications or digital magnetic recording systems [13]. In this paper, we adopted the KRLS algorithm to realize the online prediction of the health condition of the semiconductor equipment and based on the prediction, a classification procedure is used to classify the degrading stage of the equipment.

## II. PROBLEM FORMULATION

### *A. Background*

In the etching process of a semiconductor manufacturing, the contamination level of the chambers in an etching tool is measured daily by checking the particles on a testing wafer. It is called particle failure when the particle counts on testing wafer exceed a predefined threshold as it indicate the contamination level in the chamber is high and the quality of wafer produced will be affected. To restore the chamber to good condition, wet-clean maintenance is conducted periodically to the etching tools to prevent the particle failure. In general wet-clean maintenance can be planed based on a fixed production time. However, due to high mix of recipes used in these tools result in unpredictable contamination level in the chamber, thus affecting wafer etching production.

The challenge is how to assess the contamination level so that the wet-clean maintenance can be carried out before particle failure occurs. To design an effective approach to estimate the etching tool health condition, the both etching process data with recipe information and inspection data (e.g. particle counts) can be used.

*B. Degrading Stages & Prediction Accuracy Definitions*

The etching tool conditions can be categoried into 4 stages based on the particle measurement. Stage one represents a health condition of equipment and Stage 4 represents a severe degraded condition as particle failure. Stage Two and Stage Three are the intermediate stages between these two extreme stages. The Degraded Stages for measurement is defined as following:

TABLE I DEGRADING STAGES DEFINITION

| Contamination measurement |
|---|
| • Stage 1: <= R1 |
| • Stage 2: R1– R2 |
| • Stage 3: R2 – R3 |
| • Stage 4: >= R3, where R1, R2 and R3 are integer |

R1, R2, R3 are integers that can be defined based on the classification of contamination levels. According to industry filed engineer, it is known that when the measurement reaches Stage 4, a failure will occur.

*C. Recipe Based Batch Processing Data*

The production data for etching process can be easily obtained. It includes the recipes used for each batch process and the process time. The recipe contains the specification for gas flow rates, RF power and times period for each wafer processing step in the chamber of a etching tool. The recipe based batch processing data contains process information about wafer material, gases usage and RF power that affect the contamination of the chamber.

The problem of predict the contamination level is to build a learning model from historical data of the stage Y of the chamber with the recipe based batch processing data X. The model can then be used for estimating etching tool condition given online batch processing data. The main features in data X include the 8 gases' volume and RF power used in the daily production.

## III. KERNEL RECURSIVE LEAST-SQUARES (KRLS) ALGORITHM

The KRLS algorithm proposed in [2] is adopt for build the online learning model for predicting health condition of the etching tool. It performs linear regression in a high-dimensional feature space induced by a Mercer Kernel and recursively construct minimum-mean-squared-error solutions to nonlinear least-squares problems. To make the algorithm computational efficient, a sequential sparsification process is designed to curtail the growth of the structure.

*A. Learning Problem Setting*

Suppose $\mathcal{H}$ stands for a Hilbert Space and $\langle \bullet,\bullet \rangle_{\mathcal{H}}$ is the inner product in the Hilbert Space. From Mercer's Theorem [2], it follows that there exists a mapping $\phi : \chi \rightarrow \mathcal{H}$ such that $k(x,x') = \langle \phi(x), \phi(x') \rangle$ where $k(x,x')$ is a positive definite and symmetric kernel function. A commonly used kernel is the Gaussian kernel

$$\phi(x_t)^T \phi(x_i) = k(x_t, x_i) = \exp(-\|x_t - x_i\|^2 / 2\sigma^2)$$

1

*where* $i = 1, \ldots m.$

Given a recorded set of inputs and outputs

$$Z^t = \{(x_1, y_1), \ldots, (x_t, y_t)\} \qquad (2)$$

The goal is to find the best predictor $\hat{y}_{t+1}$ for $y_{t+1}$ based on $Z^t \cup \{x_{t+1}\}$. The outputs using kernel methods are nonparametric in nature and are typically of the form

$$\hat{y}_t = \sum_{i=1}^{t-1} \alpha_i \langle \phi(\tilde{x}_i), \phi(x_t) \rangle = \sum_{i=1}^{t-1} \alpha_i k(\tilde{x}_i, x_t) \qquad (3)$$

The dictionary $D_{t-1} = \{\tilde{x}_j\}_{j=1}^{m_{t-1}}$ is the kernel centers which are selected from the training samples $\{x_i\}_{i=1}^{t}$, where $m_t$ is the size of the dictionary. To find the best predictor, we may try to minimize at each time step $t$ the sum of the squared errors

$$R(\boldsymbol{\alpha}) = \sum_{i=1}^{t} (\hat{y}_i - y_i)^2 = \| \mathbf{K}_t \boldsymbol{\alpha} - y_t \|^2 . \qquad (4)$$

$$e_t = \hat{y}_i - y_i . \qquad (5)$$

where $[\mathbf{K}_t]_{i,j} = k(x_i, x_j)$ with $i, j = 1, \ldots, t$ is the full kernel matrix.

*B. Online Sparsification*

Prediction of the tool health condition is an online process which requires the computational efficiency of the learning algorithm. A common concern of kernel-based methods is that the computational complexity explodes with dimensionality. Therefore, the learning process needs regularization which adapt the structure over time as part of the learning process. KRLS involves a sparcification rule based on the approximate linear dependence (ALD) condition to curtail the growth of the kernels. Instead of making all the inputs as centers, only a subset of the points $x_1 \ldots x_t$ is selected as kernel centers in the calculation of $\hat{y}_t$ in (3). The selection procedure based on the ALD condition is as follows:

$$\delta_t = k_{tt} - \tilde{\mathbf{k}}_{t-1}(x_t)^T \boldsymbol{\alpha}_t \le v \tag{6}$$

where $k_{tt} = k(x_t, x_t)$ and $(\tilde{\mathbf{k}}_{t-1}(x_t))_i = k(\tilde{x}_i, x_t)$.

In an online scenario, at each time step t, we will face two possibilities.

*1)* If $\delta_t > v$, then a new kernel center is added into the dictionary $D_t = D_{t-1} \cup \{x_t\}$ and $m_t = m_{t-1} + 1$.

*2)* If $\delta_t \le v$, then $D_t = D_{t-1}$ and $m_t = m_{t-1}$.

### C. *Weight Update*

The KRLS update equations are derived for each of these cases in Section B.

1) *Initialize:* $v$, $\tilde{\mathbf{K}}_1 = [k_{11}]$, $\tilde{\boldsymbol{\alpha}}_1 = (y_1 / k_{11})$, $\mathbf{P}_1 = [1]$, $m_1 = 1$.

$$\tilde{\boldsymbol{\alpha}}_t = \begin{bmatrix} \tilde{\boldsymbol{\alpha}}_{t-1} - \dfrac{\boldsymbol{\alpha}_t}{\delta_t}(y_t - \tilde{\mathbf{k}}_{t-1}(x_t)^T \tilde{\boldsymbol{\alpha}}_{t-1}) \\ \dfrac{1}{\delta_t}(y_t - \tilde{\mathbf{k}}_{t-1}(x_t)^T \tilde{\boldsymbol{\alpha}}_{t-1}) \end{bmatrix}$$

2) *Begin Cycle:*

**For** $t = 2, 3 \ldots$

a) Get new sample $(x_t, y_t)$

b) Compute $\tilde{\mathbf{k}}_{t-1}(x_t)$

c) ALD test:

$$\boldsymbol{\alpha}_t = \tilde{\mathbf{K}}_{t-1}^{-1} \tilde{\mathbf{k}}_{t-1}(x_t)$$

$$\delta_t = k_{tt} - \tilde{\mathbf{k}}_{t-1}(x_t)^T \boldsymbol{\alpha}_t$$

**if** $\delta_t > v$, $D_t = D_{t-1} \cup \{x_t\}$

Compute

$$\tilde{\mathbf{K}}_t^{-1} = \frac{1}{\delta_t} \begin{bmatrix} \delta_t \tilde{\mathbf{K}}_{t-1}^{-1} + \boldsymbol{\alpha}_t \boldsymbol{\alpha}_t^T & -\boldsymbol{\alpha}_t \\ -\boldsymbol{\alpha}_t^T & 1 \end{bmatrix}$$

$$\mathbf{P}_t = \begin{bmatrix} \mathbf{P}_t & 0 \\ 0^T & 1 \end{bmatrix}$$

$$\tilde{\boldsymbol{\alpha}}_t = \begin{bmatrix} \tilde{\boldsymbol{\alpha}}_{t-1} - \dfrac{\boldsymbol{\alpha}_t}{\delta_t}(y_t - \tilde{\mathbf{k}}_{t-1}(x_t)^T \tilde{\boldsymbol{\alpha}}_{t-1}) \\ \dfrac{1}{\delta_t}(y_t - \tilde{\mathbf{k}}_{t-1}(x_t)^T \tilde{\boldsymbol{\alpha}}_{t-1}) \end{bmatrix}$$

$$m_t = m_{t-1} + 1$$

**else** $D_t = D_t$

$$\mathbf{q}_t = \frac{\mathbf{P}_{t-1}\boldsymbol{\alpha}_t}{1 + \boldsymbol{\alpha}_t^T \mathbf{P}_{t-1} \boldsymbol{\alpha}_t}$$

Compute $\mathbf{P}_t = \mathbf{P}_{t-1} - \dfrac{\mathbf{P}_{t-1}\boldsymbol{\alpha}_t \boldsymbol{\alpha}_t^T \mathbf{P}_{t-1}}{1 + \boldsymbol{\alpha}_t^T \mathbf{P}_{t-1} \boldsymbol{\alpha}_t}$

$$\tilde{\boldsymbol{\alpha}}_t = \tilde{\boldsymbol{\alpha}}_{t-1} + \tilde{\mathbf{K}}_t^{-1} \mathbf{q}_t (y_t - \tilde{\mathbf{k}}_{t-1}(x_t)^T \tilde{\boldsymbol{\alpha}}_{t-1})$$

end if

end for

Output: $D_t$, $\tilde{\boldsymbol{\alpha}}_t$.

### D. *Proposal of Online Predictive Maintenance Approach*

Based on KRLS algorithm, we propose an online predictive maintenance approach as show in Fig. 1. We can see from Fig. 1 that if we know the actual value of the indicator, we can classify the degradation stage by a simple threshold classification according to the definition in Table I. and the maintenance engineers can decide when to conduct the wet-cleaning task of maintenance based on the predicted health condition. In such cases, accuracy is the most important requirement for the learning algorithm.

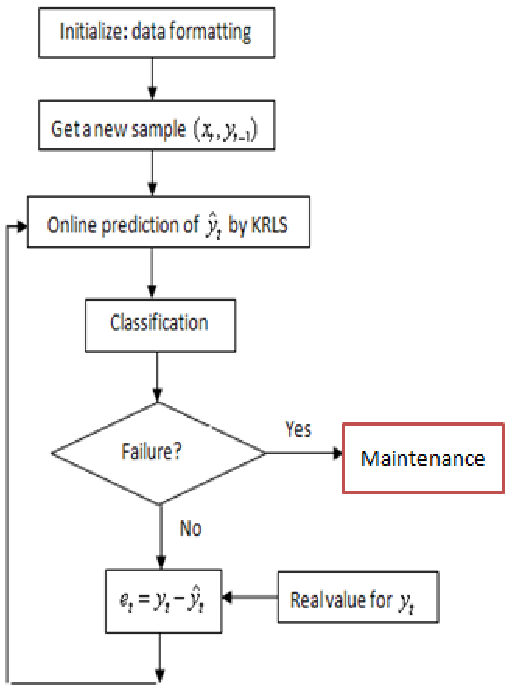


Figure 1: Online Predictive Maintenance Approach

## IV. EXPERIMENTAL TESTING

The proposed online predictive maintenance approach is tested using the dataset from an etching tool in a semiconductor company. The input data include parameters collected during production run and each set corresponds to one measurement of contamination level of the etching tool. There are total of 150 of input-output time-series data point with irregular time interval. The KRLS training algorithm is used for online training of the data set from the semiconductor company. The dimension of input vector is 10. To minimize irregular time interval impacts on the accuracy of the prediction, the time interval is also being considered as an inputs parameter. The output of the data set is the measurement of contamination level which is an indicator for the equipment health condition and classified with 4 stages in Table I.

Mean square error is a risk function which is used to quantify the difference between value implied by an estimator and true value. It is commonly used to measure the error in an estimation process. If $\hat{Y}$ is the output of n predictions, while Y is the true output, the mean square error can be written as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2$$

This formula is implemented in the system.

## V. PARAMETER OPTIMIZATION

From the theoretical analysis in Part B and Part C of Section III, it is obvious that two parameters affect the learning performance. One is the threshold value $\nu$ called accuracy parameter for online sparcification and the other is the kernel width parameter $\sigma$. During the experiment, it is observed that if the kernel width is fixed, we can find a suitable $\nu$ to obtain the smallest MSE as shown in Figure 2.

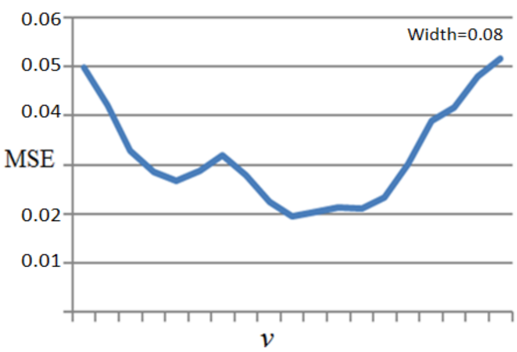


Figure 2: MSE varies with v, w=0.08

In most literatures on KRLS, the accuracy parameter $\nu$ is a fixed value and is predefined by authors. As there is no valid method provided to get the suitable width and accuracy parameter, the relationship between width and accuracy parameter must be understood: are they affecting each other in the selection method? Theoretically, width is not related with the accuracy parameter $\nu$ , as width mainly depends on the data distribution. Experiments had demonstrated that the width is independent from the threshold. Optimal combination of width and accuracy parameter can be obtained using ten-cross validation method with different values assigned to accuracy parameter. The experimental result is shown in the table II.

TABLE II PARAMETER SELECTION

| Accuracy parameter $\nu$ | Optimal Width |
|---|---|
| 0.5 | 0.2 |
| 0.1 | 0.2 |
| 0.05 | 0.3 |
| 0.01 | 0.4 |
| 0.005 | 0.5 |
| 0.001 | 0.5 |
| 0.0005 | 0.5 |
| 0.0001 | 0.7 |
| 0.00005 | 0.6 |
| 0.00001 | 0.8 |

From Table II, we can find that width is varying while accuracy parameter changes. They are affecting each other. This is mainly because in the application kernel method, the accuracy is predefined parameter to run the algorithm and find the optimal width, which is generating the smallest mean square error. Thus, it is important to find an optimal combination of accuracy parameter and kernel width.

TABLE III THE EXPERIMENTAL RESULTS

| Method | Time consumption | MSE | Parameters |
|---|---|---|---|
| Original | 3 seconds | $3.60\,e^{-4}$ | Width = 0.5, v=0.005 |
| Ten-Cross validation | 568 seconds | $9.25\,e^{-5}$ | Width = 0.5, v=0.001 |

Based on the above observation, we proposed a method to obtain an optimal combination of width and accuracy parameters. Using the optimal set of parameters can made great improvement on the accuracy, though it consumes a lot of computing time to cover a large range of both parameters. This method is mainly making changes with ten-cross validation. It consists of the following steps:

1. Provide a series of possible numbers for accuracy parameter, and initialize MSE as positive infinity.
2. Pick a number from the series as accuracy parameter.
3. From width=0.1 to width=0.9, conduct ten cross validation.
4. Compute the mean square error for each result and compare it with MSE, update the smaller value as

MSE and record the width and accuracy parameter that generates MSE.

5. Remove the accuracy parameter from the series, if the series is not empty, repeat step 2; if the series is empty, return the width and accuracy parameter that can generate MSE. This is the optimal combination.

In the current implementation, the series of possible number for accuracy parameter includes $\{8e^{-3}, 5e^{-3}, e^{-3}, 8e^{-4}, 5e^{-4}, e^{-4}, 8e^{-5}\}$. Table III shows the comparison results by using one of the parameter combinations in Table II and our proposed ten-cross validation methods. It is obvious that ten-cross validation can obtain smaller MSE in sacrifice of the time consumption.

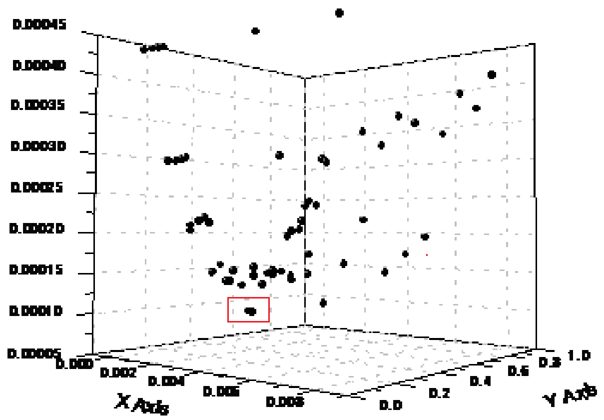


Figure 3: MSE over Accuracy Parameter and Width

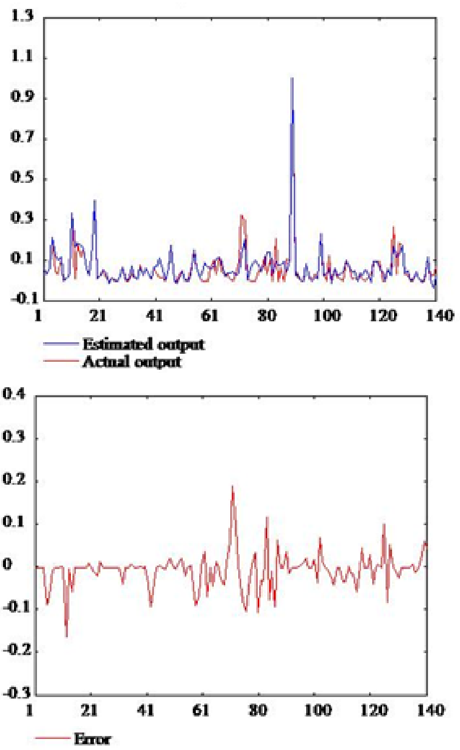


Figure 4: Prediction outputs and its MES

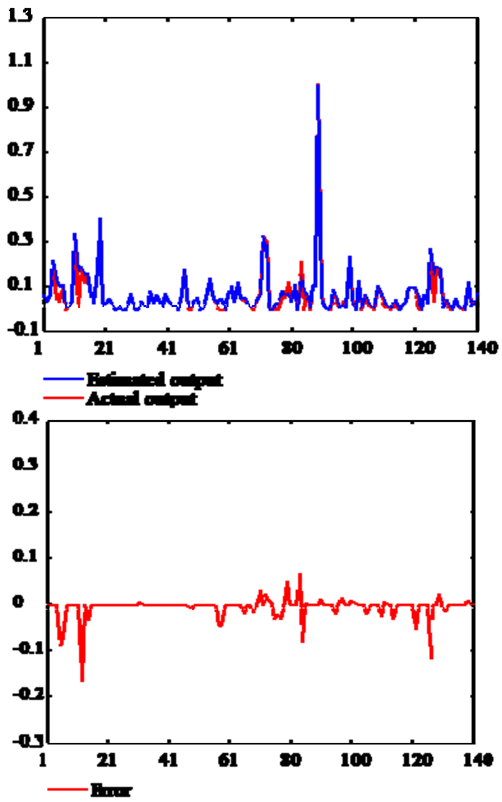


Figure 5: Prediction outputs and its MES with optimal parameter set

A 3-D graph of MSE for different combinations of accuracy and width parameters is generated as shown in Figure 3. The X axis represents the accuracy parameter, the Y axis represents the width, while the Z Axis represents the MSE. From Figure 3. it is easy to find the optimal combination of parameter (indicated by red rectangle). Figure 4 shows the KRLS prediction output using parameter combinations in Table II, and its MES when compared with the true outputs. Figure 5 shows the KRLS prediction output using the optimal parameter set generated from the ten-cross validation based optimization method. It is obvious that running KRLS with optimal parameter set can reduce prediction error. It is observed from Figure 5 that most of the estimation errors lie near zeros which means the prediction is more accurate. Even though there are relatively large errors at some time stamps, the learning system can still predict the spike. The outputs from the learning system can be classified to the degrading stages of the equipment accurately.

## VI. CONCLUSIONS

An online predictive maintenance approach based on KRLS learning is proposed to assess health condition for semiconductor equipment. The KRLS algorithm can be used with a relatively small size training dataset for online prediction. The computational efficient is achieved through the involvement of the sparcification procedure. Based on the prediction of contamination measurement, threshold classification can be used to determine which stage the equipment is in. If the prediction indicates it is in stage 4, the equipment needs maintenance. Simulation results show that the proposed online predictive maintenance approach can reach higher accuracy in terms of 4 stages of health indicator prediction and be computational efficiency.

A ten-cross validation based parameter optimization approach is used for improving the accuracy of the online prediction. With the optimal set of the parameters MSE can be reduced by 10 times though it consumes more computing time. We would only recommend conducting offline optimization for parameters adjustment. Further study should be carried out for an effective online optimization to obtain optimal set of parameter with less computation time.

## REFERENCES


[1] Y. Segawa, M. Ohnishi, and T. Ibaraki, "Optimal minimal-repair and replacement problem with age dependent cost structure" in Computers & Maintenances with Applications, Oxford, 24, 91-101.: 10, 1016/0898-1221(92)90233-8. , 1992.

[2] Y. Engel, S. Mannor and R. Meir, "The kernel recursive least-squares algorithm," IEEE Trans. Signal Processing, vol. 52, no. 8, pp. 2275 -2285, Aug. 2004.

[3] W. Liu, P. P. Pokharel and J. C. Principe, "The kernel least-mean-square algorithm," IEEE Trans. Signal Processing, vol. 56, no. 2, pp. 543 - 554, Aug. 2008.

[4] W. Liu and J. C. Principe, "Kernel affine projection algorithms," EURASIP J. Adv. Signal Process, Aug. 2008.
W. Liu and J. C. Principe, "Sliding window generalized kernel affine projection algorithm using projection mappings," EURASIP J. Adv. Signal Process, vol. 16, 2008.

[5] W. Liu, Il. Park, Y. Wang and J. C. Principe, "Extended Kernel Recursive Least Squares Algorithm," IEEE Trans. Signal Processing, vol. 57, no.10, pp. 3801 - 3814, Oct. 2009.

[7] Jemwa, G.T. and Aldrich, C., "Kernel-based fault diagnosis on mineral processing plants" Minerals engineering, vol. 19, no. 11, pp. 1149 - 1162, 2006.

[8] Chattopadhyay, A. and Das, S. and Coelho, C.K., "Damage diagnosis using a kernel-based method," Insight-Non-Destructive Testing and Condition Monitoring, vol. 49, no. 8, pp. 451 - 458, 2007.

[9] Desforges, M. J. and Jacob, P. J. and Ball, A. D. , "Fault detection in rotating machinery using kernel-based probability density estimation," International Journal of Systems Science, vol. 31, no.11, pp. 1411-1426, 2000.

[10] Yin, Y. J. and Zhan, J. P. and Guo, C. X. and Wu, Q. H. and Zhang, J. M., "Multi-kernel support vector classifier for fault diagnosis of transformers," 2011 IEEE Power and Energy Society General Meeting, pp. 1 - 7, 2011.

[11] Van Calster, B. and Timmerman, D. and Lu, C. and Suykens, J. A. K. and Valentin, L. and Van Holsbeke, C. and Amant, F. and Vergote, I. and Van Huffel, S., "Preoperative diagnosis of ovarian tumors using Bayesian kernel-based methods," 2011 Ultrasound in obstetrics and gynecology, vol. 29, no.5, pp. 496-504, 2007.

[12] Ahmed, T. and Coates, M. and Lakhina, A., "Multivariate online anomaly detection using kernel recursive least squares" in Proceedings of 26th IEEE International Conference on Computer Communications, 625-633, 2007.

[13] Vaerenbergh, S. and Via, J. and Santamana, I. , "Van A sliding-window kernel RLS algorithm and its application to nonlinear channel identification," in Proceedings of 2006 IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 5, pp. V – V, 2006.

[14] Analyzing APS Models Statistically httm://www.gepsoft.com/gepsoft/aps3kb/Chapter09/Section2/SS03.htm

[15] Chen, C., "Signal and image processing for remote sensing" CRC Press Technology & Industrial Arts 648 pages ISBN 0849350913, 2006