IDFD-O-013

Anomaly Detection Using Average Normal Profile of Time-Series Process Data

Toru KASHIMA
kashima-toru@jp.yamatake.com
Yamatake Corporation
1-12-2 Kawana Fujisawa Kanagawa JAPAN 251-8522
Phone: +81 -466-20-2430 Fax: +81-466-20-2430

Abstract:
This paper presents a method for calculating an "average normal profile" from plural sets of time-series process data, coping with fluctuations of data in the time dimension, differences in the numbers of data points, and mismatches of sampling time. DTW (Dynamic Time Warping [1]) is sequentially used in conjunction with the weighted average to calculate the profile in this method. After obtaining the profile, the difference between the profile and a set of time-series data is calculated as a "distance" which is used to detect anomalies. The method is applied to simulated data and actual CVD process data and the results show that the method can detect anomalies.

Motivation:
Statistical values, such as mean, maximum, minimum or standard deviation values, are ordinarily used for SPC (Statistical Process Control). These values, however, are often useless for detecting anomalies in the practical manufacturing process (for example, when the process data naturally fluctuates in the time dimension). Figures 1 and 2 show such an example using simulated data. Figure 1 shows the forward power of the dry etching process in a recipe step where the power rapidly rises from around 0 to around 200. Twenty sets of normal wafer data is shown in (a) and one set of abnormal wafer data is shown in (b). The rising timing is about 0.2 second after the beginning, but the timing fluctuates even among the normal wafers. A downward spike noise occurs just after the rising in the abnormal data. Figure 2(a), (b), (c) and (d) respectively show the mean, maximum, minimum and standard deviation values of all wafers. The rightmost triangular point in each graph is calculated from the abnormal data. It is impossible to detect the anomaly only using these values.

Approach:
The presented method can cope with this problem using DTW. DTW is ordinarily used to calculate the difference between two sets of time-series data, coping with fluctuations of data in the time dimension, differences in the numbers of data points, and mismatches of sampling time. The difference represents the total cost that accrues while following "the warping path", shown in Figure 3, which is calculated by DP (Dynamic Programming) matching. The more different the profiles of data are, the longer the distance becomes.

Figure 4 shows the procedure of the presented method. The method obtains the warping path by executing DTW, averages data at each data point along with the path, regards the average data as a new set of data, and iterates over it. The influence of the order in which a set of data is selected to execute DTW can be eliminated by using the weighted average. Once the average normal profile is obtained, the distances to the profile can be calculated by DTW.

Application Results:
The average normal profile was obtained from twenty normal wafers of the simulated data mentioned above and is shown in Figure 5(a). The distances to the profile are shown in Figure 5(b). The anomaly can be detected because the rightmost triangular point that is calculated from the abnormal data shows the largest value.

Figure 6(a) shows actual data (the plate heater current of a CVD tool) of 74 wafers. The wafer thickness uniformity degenerated as the wafer transfer system of the tool had trouble locating a wafer on the plate. The uniformity improved after maintenance of the transfer system. Figure 6(b) shows the values of uniformity of 62 wafers before maintenance and 12 wafers after maintenance. Figure 7(a) shows the average normal profile calculated from the latter 12 wafers' data by the presented method. Figure 7(b) shows the distances to the average normal profile. The distances after maintenance are smaller than the distances before maintenance as well as the values of uniformity. The correlation coefficient between the values of uniformity and the distances is 0.68.

Conclusions:
The presented method was applied to the simulated data and the actual process data. The results indicate that the method can detect anomalies which can not be detected by the ordinary statistical values and can correlate time-series process data to wafer metrology data.

Acknowledgement:
This work was supported by Eiichi Murayama and Yasuhito Tanaka from Sony Corporation. They provided the author with not only the actual process data but also opportunities for valuable discussion.

References:
[1] Yamada et al, "Classification by Time-series Decision Tree", The 17th Annual Conference of the Japanese Society for Artificial Intelligence, 2003.

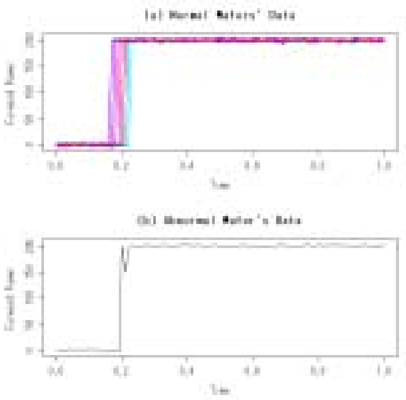

Figure 1. Time-Series Data of Forward Power

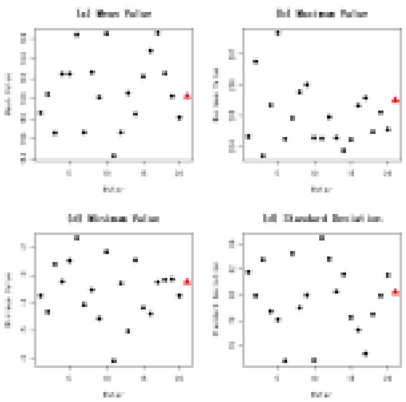

Figure 2. Statistical Values

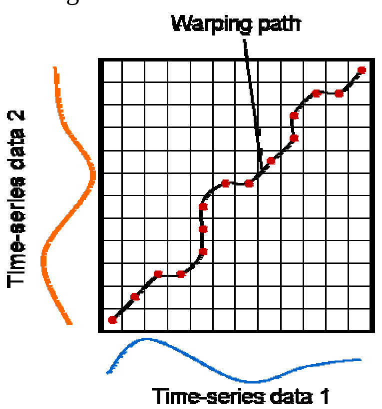

Figure 3. Dynamic Time Warping

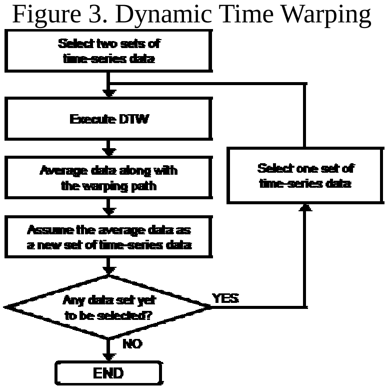

Figure 4. Flow Chart of the Procedure

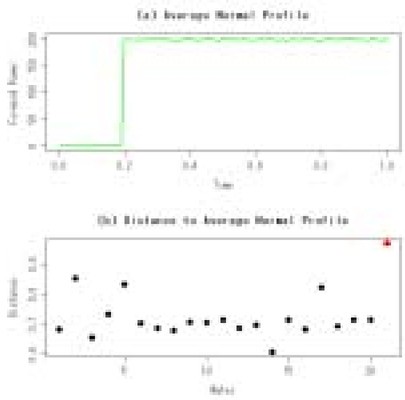

Figure 5. Average Normal Profile and Distance

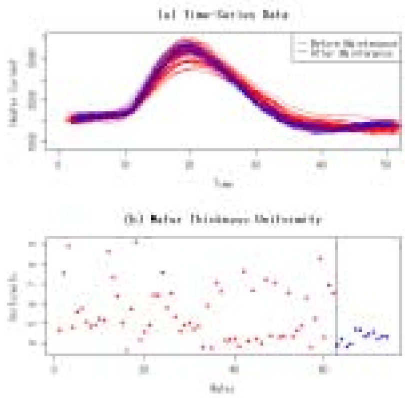

Figure 6. Heater Current and Uniformity

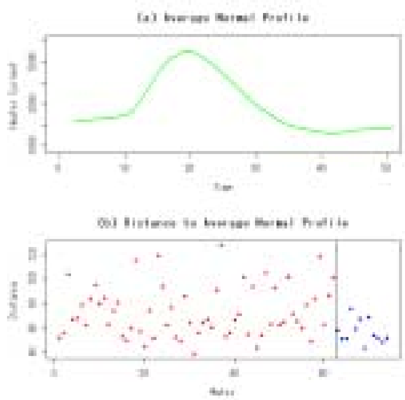

Figure 7. Average Normal Profile and Distance