Expert Systems with Applications 41 (2014) 2619–2629

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

# Data based segmentation and summarization for sensor data in semiconductor manufacturing

Eunjeong L. Park [a], Jooseoung Park [a], Jiwon Yang [a], Sungzoon Cho [a,*], Young-Hak Lee [b], Hae-Sang Park [b]

[a] Department of Industrial Engineering, Seoul National University, 1 Gwanakro, Gwanakgu, 151-744, Seoul, Republic of Korea
[b] Samsung Electronics Co. Ltd., 416 Maetan-dong, Yeongtong-gu, Suwon, Gyeonggi-do, Republic of Korea

ARTICLE INFO

*Keywords:*
Semiconductors
Sensor data
Segmentation
Summarization
Free knot spline with knot removal
Time series sensor segmentation

ABSTRACT

In semiconductor manufacturing processes, sensor data are segmented and summarized in order to reduce storage space. This is conventionally done by segmenting the data based on predefined chamber step information and calculating statistics within the segments. However, segmentation via chamber steps often do not coincide with actual change points in data, which results in suboptimal summarization. This paper proposes a novel framework using abnormal difference and free knot spline with knot removal, to detect actual data change points and summarize on them. Preliminary experiments demonstrate that the proposed algorithm handles arbitrarily shaped data in a robust fashion and shows better performance than chamber step based segmentation and summarization. An evaluation metric based on linearity and parsimony is also proposed.



## 1. Introduction

In semiconductor manufacturing processes, sensor data are recorded on a second by second basis. Sometimes referred to as trace data, they are a sequence of discrete time signals, heavily used for supervisory purposes such analytics or monitoring. Examples of analytics include yield prediction and management, fault detection and classification (FDC) (He & Wang, 2007), virtual metrology (Kang et al., 2009), while monitoring refers keeping an eye on manufacturing parameters' trends (Qin, Cherry, Good, Wang, & Harrison, 2006; Sarfaty et al., 2002; Spanos, Guo, Miller, & Levine-Parrill, 1992). The implementation of such supervisory tasks allows automatic maintenance as well as effective quality control, and thus benefit a semiconductor fabrication plant with dramatic savings in cost. Therefore the appropriate utilization of sensor data is crucial for a semiconductor company, as the industry continuously faces price competitions.

A semiconductor company's sensor data amounts up to units of gigabytes per day. Therefore they are compressed into smaller and manageable bits of information that are easier to access and analyze, in spite of the information loss. This task is normally carried out in two steps: *segmentation*, where data are segmented whenever the data trend has changed, and *summarization*, where representative values are calculated from each segment. Indeed, segmentation can be best done by manually finding change points in the data. However, thousands of wafers are recorded concurrently by myriad sensors in a semiconductor manufacturing process. Finding a subjective set of change points among such a big data is unrealistic or otherwise costly without computational means.

A frequently used solution addressing this problem employs the manufacturing recipe. In this particular solution, each segment is set equal to the interval between the beginning and end of each *chamber step*, a set of instructions in the manufacturing recipe. Once the segments are defined, each segment is summarized by statistics, such as the minimum, maximum, mean, and standard deviation.

The assumption behind chamber step based segmentation is that a data change point will occur if and only if the chamber step changes. But this assumption is not always correct. Take Fig. 1(a) for instance, where chamber step based segmentation partitions the data into seven segments. Ideally, no segmentation is needed for $T_1$, because the underlying signal is invariant. On the other hand, $T_2$ needs to be segmented with much finer granularity, so that each segment can represent a unique trend in data. Sometimes segmentation points do not coincide with actual change points as in Fig. 1(b), because manufacturing actions often show latencies before actual data changes take place.

The other assumption of chamber step based segmentation is that the minimum, maximum, mean, and standard deviation are sufficient for representing a segment. It is satisfied if the data are *discrete*, where the sensor data values are constant within each segment as in Fig. 1(a). The minimum and maximum reflect the

* Corresponding author. Tel.: +82 01 7203 6275.
*E-mail addresses:* ejpark04@snu.ac.kr (E.L. Park), uni208@snu.ac.kr (J. Park), ad1392@snu.ac.kr (J. Yang), zoon@snu.ac.kr (S. Cho), younghaklsi.lee@samsung.com (Y.-H. Lee), haes.park@samsung.com (H.-S. Park).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.11.001

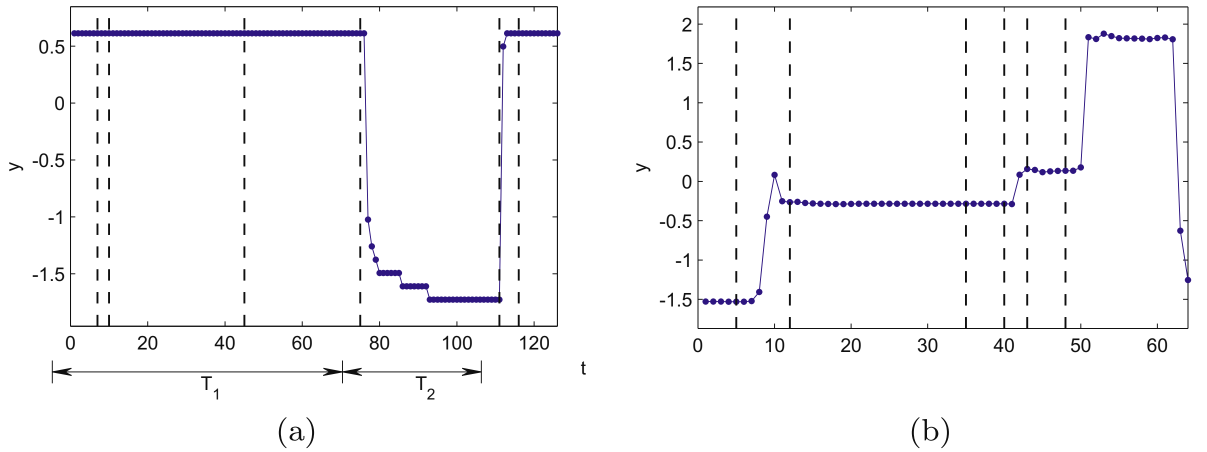


**Fig. 1.** Two example cases where chamber step based segmentation fails to match actual signal changes. Sensor data values ($y$-axis) are plotted along time ($x$-axis), and dotted lines represent chamber step changes. $T_1$ in (a) is an example where data is segmented though the signals did not change, whereas $T_2$ was not segmented while the signals change. In (b), chamber steps do not coincide with actual change points.

outliers of a segment while the mean and standard deviation represent the central tendency and dispersion from the mean value, respectively. However, for *continuous*-valued data, where diagonally aligned data exist, this assumption may not be appropriate. Consider a simple situation where values monotonically increase. The minimum and maximum value in this case would be the first and last data point of the data, respectively. The mean would be no more than the average value between the minimum and maximum, and the standard deviation would not represent the residual sum of squares but the total sum of squares in terms of linear regression.

Then what are the conditions of a 'good' segmentation and summarization? We propose two conditions for a good segmentation in this paper: *linearity* and *parsimony*. 'Good segments' are acquired based on these criteria, and 'good summarizations' are attained by calculating valid, adequate representative values from each segment.

Much research has been carried out in various fields to approach the task of segmentation: Kalman filters in signal processing, change point detection algorithms in machine learning, free knot spline with knot removal in applied mathematics. However, a common problem with manufacturing data is that the data vary in shapes, and none of the shapes is dominant Fig. 2. Hybrid dynamical systems such as shifting Kalman filters have been introduced in the literature to cover this issue, but are computationally burdensome in real world situations.

In this paper, a robust segmentation algorithm based on linearity and parsimony and a corresponding summarization framework are proposed. The segmentation algorithm in particular, assumes that sensor data either has a discrete or continuous pattern. If the data is considered discrete, it is segmented piecewise constantly, and if continuous, it is segmented piecewise linearly. In the summarization phase, representative values are calculated based on the characteristics of the derived segments. The whole framework covering segmentation and summarization is aimed to be as computationally light as possible, in order to be practical in real world situations.

The remainder of this paper is organized as follows: Section 2 presents the related works of segmentation and summarization. Section 3 presents a preprocessing method of aggregating data of many wafers achieved from each sensor. Section 4 addresses the design of a piecewise linear segmentation framework for sequential data. Section 5 discusses an approach to effectively summarize the data with the given segments derived above. In Section 6, experiments of the given technique is performed on real semiconductor manufacturing data. Finally, a conclusion is stated in Section 7.

## 2. Related work

In areas such as signal processing, mathematical statistics, automatic control, communication systems and quality control, data segmentation has various names such as change point detection (CPD) and signal segmentation.

In the applied mathematics discipline, free knot spline with knot removal has been applied for data segmentation. Free knot spline with knot removal mainly consists of two parts: free knot spline and knot removal. Among these, the free knot spline part was introduced by De Boor (1974), De Boor and Rice (1968), and further developed by scholars such as Jupp (1978) or Schütze and Schwetlick (1997, 1995).

Free knot splines is a data smoothing algorithm that optimizes the positions of knots, so that the error is minimized. There are two important model parameters in this algorithm, namely the number of knots and the order of spline functions $r$. In order to automatically optimize the number of knots, knot removal was proposed by Lyche, Cohen, and Mørken (1985) and Lyche and Mørken (1987). Therefore, combining the two algorithms of free knot spline and knot removal, data segmentation is conducted by only determining one model parameter – the order of spline functions $r$.

$$RSS(f,\kappa_{FS}) = \sum (y_i, f(t_i))^2 + \kappa_{FS} \int \{f^{(r+1)}(p)\}^2 dp \qquad (1)$$

Due to its nature in smoothing the data, the free knot spline with knot removal algorithm gives good segmentation results with continuous-valued data. However, it fails to find the exact change points for discrete-valued data, due to the same nature. Moreover, iterative optimization and knot removal are both time consuming processes.

A number of change point detection algorithms based on the Bayesian theory were also proposed. In particular, a change point detection method which performs Bayesian curve fitting using Markov chain Monte Carlo (MCMC) was proposed by Fearnhead and Liu (2007) and Punskaya, Andrieu, Doucet, and Fitzgerald (2002). While these papers are trained in batches, Adams and MacKay (2007) has proposed a probability-based online change point detection method, and Saatçi, Turner, and Rasmussen