Computers & Industrial Engineering 85 (2015) 132–144

Contents lists available at ScienceDirect

Computers & Industrial Engineering

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

Phase I analysis of multivariate profiles based on regression adjustment ☆

Jiajia Zhang [a], Haojie Ren [a,*], Rui Yao [b], Changliang Zou [a], Zhaojun Wang [a]

[a] Institute of Statistics, Nankai University, Tianjin, China
[b] Beijing Institute of Control Engineering, Beijing, China

ARTICLE INFO

Article history:
Received 3 November 2014
Received in revised form 23 February 2015
Accepted 25 February 2015
Available online 19 March 2015

Keywords:
Functional principal component analysis
Nonlinear profile
Nonparametric regression
Profile analysis
Statistical process control

ABSTRACT

The use of statistical process control in monitoring and diagnosis of process and product quality profiles remains an important problem in various manufacturing industries. Although the analysis of profile data has been extensively studied in the literature, the challenges associated with monitoring and diagnosis of multiple functional profiles are yet to be well addressed because it is usually difficult to properly model the inter-relationship of multiple profiles. Motivated by a real-data application in semiconductor industries, we develop a new modelling and monitoring framework for Phase-I analysis of multiple profiles. The proposed framework incorporates the regression-adjustment technique into the functional principal component analysis. In this framework, the multiple profiles are treated as multivariate functional observations and their regression-adjusted residuals are used for monitoring. Simulation results show that the proposed method could describe the major structure of profile variation well and effectively find outlying profiles in a historical dataset due to sufficiently utilizing the information on between-profiles correlation.

1. Introduction

Because of recent progress in sensing and information technologies, automatic data acquisition has become the norm in various industries. Consequently, a large amount of quality-related data of certain processes have become available. Statistical process control (SPC) based on such data is an important component of process monitoring and control. In many applications, the quality of a process is characterized by the relationship between a response variable and one or more explanatory variables. A collection of data points of these variables can be observed at each sampling stage, which can be represented by a curve (i.e., profile). In some calibration applications, the profile can be described adequately by a linear regression model. In other applications, however, more flexible models are necessary in order to describe the profiles properly. An extensive discussion of the related research problems has been given by Woodall, Spitzner, Montgomery, and Gupta (2004).

Profile monitoring has been extensively studied in SPC and several methods have been developed for monitoring linear and nonlinear profile data. Some examples include the use of multivariate control charts for monitoring linear and nonlinear regression coefficients (Kang & Albin, 2000; Mahmoud & Woodall, 2004; Zou, Tsung, & Wang, 2007; Williams, Woodall, & Birch, 2007), monitoring methods based on mixed-effect models (Jensen, Birch, & Woodall, 2008; Paynabar, Jin, Agapiou, & Deeds, 2012), dimension-reduction techniques (Lada, Lu, & Wilson, 2002; Ding, Zeng, & Zhou, 2006; Chicken, Pignatiello, & Simpson, 2009; Paynabar & Jin, 2011; Viveros-Aguilera, Steiner, & Mackay, 2014), and methods for monitoring roundness profiles (Colosimo & Pacella, 2007; Colosimo, Cicorella, Pacella, & Blaco, 2014). Extensive discussion about various research problems on profile monitoring can be found in Woodall (2007), Noorossana, Saghaei, and Amiri (2011) and Qiu (2014, chap. 10).

Most recent studies concentrated on the situation with a univariate profile that only contains one response variable. Although such profiles can characterize various applications as described in the literature, multivariate functional profiles in which multiple response variables are involved simultaneously may be even more representative of most industrial applications in certain real world practices. When the correlation structure between quality characteristics is ignored and profiles are monitored separately, then misleading results may be expected (c.f., Lowry, Woodall, Champ, & Rigdon (1992) and Hawkins (1991) for relevant discussions). However, research on the monitoring and diagnosis of multivariate general profiles is still scanty. Noorossana, Eyvazian, Amiri, and Mahmoud (2010) discussed multivariate linear profile monitoring

☆ This manuscript was processed by Area Editor Min Xie.
* Corresponding author.
E-mail address: renhaojie1991@163.com (H. Ren).

http://dx.doi.org/10.1016/j.cie.2015.02.025
0360-8352/© 2015 Elsevier Ltd. All rights reserved.

in Phase I analysis, mainly based on the ordinary least square estimation. Another relevant work is Zou, Ning, and Tsung (2012) which focused on a study of the Phase II method for monitoring a general multivariate linear profile by using the LASSO-based multivariate SPC techniques. More recently, Chou, Chang, and Tsai (2014) developed a process monitoring strategy for monitoring multiple correlated nonlinear profiles. Nevertheless, how to apply conduct Phase I monitoring for general multivariate profiles including nonlinear profiles, still remains a challenge and has not been thoroughly investigated in the literature. In practice, Phase I process control is crucially important to check the stability of historical profile data and to obtain accurate estimates of the baseline model parameters used for Phase II monitoring (Zhang & Albin, 2009). The main objective of this paper is to develop a new nonparametric method for Phase I monitoring of multivariate profile data.

In the literature of nonparametric profile monitoring, the nonparametric profile model (c.f. Zou, Tsung, & Wang, 2008) is usually considered

$$y_{ij} = g(x_{ij}) + \varepsilon_{ij}, \quad j = 1, \ldots, n_i, \quad i = 1, 2, \ldots, \tag{1}$$

where $\{x_{ij}, y_{ij}\}_{j=1}^{n_i}$ is the ith sample collected over time, x_{ij} is the jth design point in the ith profile, g is a smooth nonparametric profile, and ε_{ij}s are i.i.d. normal random errors with mean 0 and variance σ^2. Zou et al. (2008) developed a procedure based on the combination of local linear smooth and traditional SPC charting techniques. Furthermore, to account for the within-profile correlations, Qiu, Zou, and Wang (2010) proposed to use the nonparametric mixed-effects model which allows flexible variance–covariance structures. Recently, Hung, Tsai, Yang, Chuang, and Tseng (2012) introduced a technique called Support Vector Regression to model the profile relationship between the response variable and explanatory variables, while the within-profile correlation is accommodated by using a resampling technique called block bootstrap. Closely related to this idea, Chuang, Hung, and Yang (2013) provided an easy-to-implement and computationally cheaper framework for monitoring nonparametric profiles by taking into account the within-profile correlation.

It is not straightforward to extend (1) to multivariate settings. One technical challenge is how to model such multivariate profiles by taking the correlations between curves into account. Recent and representative work is Soleimani and Noorossana (2014). However, such traditional methods developed for a linear regression may not fully characterize the information from a complex profile datastream where between-curves correlations may be highly related to the covariates (design points). Although certain efforts have been made on the within profile autocorrelation in simple linear profiles or parametric profiles (see Soleimani, Noorossana, & Amiri, 2009; Soleimani, Noorossana, & Niaki, 2013; Wang & Tamirat, 2014; Khedmati & Niaki, 2015), the complexity of simultaneously handling the between-curves and within-profile correlations still challenges us. Engineering applications that give rise to profile data often lead to correlated error terms. As demonstrated by Qiu et al. (2010), neglecting the within-profile correlation will result in adverse effects on both in-control (IC) and out-of-control (OC) properties of control schemes. The situation may be more serious for multivariate profile processes due to the intricacy of the model and the large number of responses. Moreover, how to integrate an appropriate regression function nonparametric test with classical SPC techniques is not quite straightforward.

In this paper, we try to deal with all the aforementioned challenges, and to resolve the latent issue of existing nonparametric modelling and monitoring methods that are unable to efficiently utilize full information from a multivariate profile process. We start by introducing a manufacturing example taken from an etching process.

2. A motivating example

We use an industrial etching process example taken from semiconductor manufacturing to illustrate the motivation for this research. The etching chamber is equipped with more than 50 sensors which record the values of several variables with time during a batch. For illustrative purposes, only 7 major variables $x_1, \ldots, x_7$ will be considered. Variable x_1 is related to spectral analysis of chamber gas, while x_2 to x_5 relate to plasma operations. Variables x_6 and x_7 are the chamber temperature and pressure, respectively. There are many steps involved in the batch operation, but here we only focus on the second step since engineers consider that this step is one of the most crucial steps which usually contains enough critical information to distinguish the out-of-control conditions in the process. The entire dataset contains 364 wafers which consist of 16 lots, while each lot represents a short-term operating cycle. We aim at using the Phase I monitoring schemes to identify any abnormal profile observations from the dataset.

The profile observations of each variable are synchronized so that the number and positions of covariate points (say, time, in this example) are equal (Akima, 1970). Fig. 1 shows the profiles of the first four profiles for 364 wafers. The x-axis represents totally 137 synchronized profile points. The profile curve of variable x_5 is quite similar to that of x_4 and thus is not presented here. The values of x_6 and x_7 for each wafer are almost identical and thus we choose to plot the sample mean of each profile against the wafer index (from 1 to 364). Clearly, the profile curves for each variable present a similar functional and smooth change along with the time point. Some of them significantly deviate from the population curves and should be regarded as outlying profiles which we would like to use some detection procedures to automatically screen out. The values of temperature x_6 and pressure x_7 vary within certain ranges and can be viewed as two environmental covariates in our following analysis.

From Fig. 1(a)–(d), we can vaguely see that there might be some interrelationships between different variables. To further confirm, we specifically present Fig. 2(a)–(c) to show the scatter plots of the raw data for the x_1 against x_3, x_2 against x_4 and x_5 against x_4, respectively, at the time point $t = 5$. It is now clear that the correlations between the variables are significant, which suggests that the profile curves have considerable interrelationships and consequently a multivariate modelling and monitoring framework is likely to be more appropriate than a univariate one.

Besides the between-curves correlations, since the measurements of all the first five variables in each profile (wafer) are taken in consecutive time intervals, the data exhibit a considerable amount of positive serial autocorrelation in each profile. For example, Fig. 3 depicts three estimated within-profile correlation curves $\hat{\rho}_j(t, t')$ against $t = 1, \ldots, 137$ for $j = 1, 3, 4$, where $\hat{\rho}_j(t, t')$ denotes the estimated correlation between the observations at t and t' time points for the jth variable. We chose $t' = 5$, 35 and 65 for illustration because similar results can also be observed for other time points (not reported here but available from the authors). We can see that within-profile correlation is substantial, and thus it should not be ignored.

Naturally, we may use the model (1) to describe the profile curves for variables x_1 to x_5. The local linear smoothing curve (Fan & Gijbels, 1996) of the average of the 364 wafers is obtained for each variable. They can be deemed as an estimate of the population profile models. Then, the corresponding residual curves are obtained by subtracting estimated population curves from the original observations. From those residual curves, some