QUALITY AND RELIABILITY ENGINEERING INTERNATIONAL
Qual. Reliab. Engng. Int. 2005; **21**:677–688
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/qre.711

Special Issue

Using Profile Monitoring Techniques for a Data-rich Environment with Huge Sample Size

Kaibo Wang and Fugee Tsung*,†
Department of Industrial Engineering and Engineering Management, Hong Kong University of Science and Technology, Kowloon, Hong Kong

In-process sensors with huge sample size are becoming popular in the modern manufacturing industry, due to the increasing complexity of processes and products and the availability of advanced sensing technology. Under such a data-rich environment, a sample with huge size usually violates the assumption of homogeneity and degrades the detection performance of a conventional control chart. Instead of charting summary statistics such as the mean and standard deviation of observations that assume homogeneity within a sample, this paper proposes charting schemes based on the quantile–quantile (Q–Q) plot and profile monitoring techniques to improve the performance. Different monitoring schemes are studied based on various shift patterns in a huge sample and compared via simulation. Guidelines are provided for applying the proposed schemes to similar industrial applications in a data-rich environment. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: statistical process control; profile; sample size; Q–Q plot

1. INTRODUCTION

Statistical process control (SPC) techniques evolved decades ago and are tailored to situations in which measurement data are very limited. With modern measurement and inspection technology, it is not uncommon to have hundreds of different process variables that are measured in huge sample sizes on short time intervals. Few researchers really look into how to make use of such high-volume data for quality control purpose. We propose to transform such a problem with huge sample size to a profile monitoring problem and develop SPC techniques that make good use of the high-volume data.

In-process sensors with huge sample size have become commonplace in the modern manufacturing industry, due to the increasing complexity of processes and products and availability of advanced sensing technology. Sensors keep track of pressure, temperature, speed, color, humidity and any other variables that are appropriate

*Correspondence to: Fugee Tsung, Department of Industrial Engineering and Engineering Management, Hong Kong University of Science and Technology, Kowloon, Hong Kong.
†E-mail: season@ust.hk

Contract/grant sponsor: RGC Competitive Earmarked Research Grants; contract/grant numbers: HKUST6183/03E and HKUST6232/04E

Received 20 April 2004
Revised 20 August 2004

to a particular process (Barry and Linoff[1]). For example, in automotive manufacturing the in-line Optical Coordinate Measurement Machine (OCMM) measures 100–150 points on each major assembly with a 100% sample rate. The OCMM, therefore, takes a tremendous amount of measurement data every 10–30 seconds and efficient use of these data becomes a challenge. Another example is the machine vision system for mobile phone display inspection, in which 5000 points on each display are measured at the rate of 10 seconds per display board. The consequence of not fully utilizing such rich data in the manufacturing process is often a ruined batch of product and a very costly shutdown and restarting of the process. Throughout this paper, we will use the mobile phone display assembly and testing process as an example to demonstrate a proposed SPC technique that is based on a profile monitoring method for such a data-rich environment.

The current practice of handling a situation with huge sample size is to apply a conventional SPC control chart to the summary statistics of each sample such as sample mean or standard deviation. However, some serious defects, such as a dark corner and a bad pixel cluster that are due to partial out-of-control data in a sample, cannot be detected by a conventional charting method, as the out-of-control data would be averaged out and buried in the huge sample.

To deal with such a unique problem with huge sample size, we suggest not monitoring the summary statistics of the sample, but characterizing the sample by a quantile–quantile (Q–Q) plot. A linear profile that fits the Q–Q plot of each sample would be monitored via a profile charting technique and the alarms of mean shifts, variance shifts and distribution shifts of partial data in the sample are expected to be triggered efficiently and effectively.

In the next section, a real example from a mobile phone assembly and testing process is used to illustrate the problem. Profile-related techniques, as a possible solution for the data-rich problem, are then reviewed. Section 2 introduces methodologies related to Q–Q plots that would characterize a huge sample and transform it to a linear profile for SPC application. Unique problems in Phase I and Phase II SPC developments with huge sample sizes and their corresponding methods are investigated in subsequent sections. The run length performances of the proposed methods are compared with different SPC monitoring schemes in Section 5. Finally, Section 6 concludes this paper and gives prospective applications of the proposed approaches.

1.1. A motivating example

A example from a mobile phone assembly and testing process is used to illustrate the problem in a data-rich environment. This is initiated by a leading electronics manufacturing services provider who aims to provide ultra-modern manufacturing technology for high-volume products. Among its contract products, the cellular phone is one for which there is a great demand for quality. In particular, the display component of the cellular phone is the most critical quality part for the end user. The display component, including a liquid crystal display (LCD), a driving board and other mechanical parts to support its movement, is assembled in one production line. After several assembling stages, the display is mounted on the testing station for quality assurance.

Traditionally, the quality assurance relies on the visual checking of defects in a display, which is inaccurate and labor intensive. These days, a machine vision system with image acquisition hardware and analysis software can perform this repetitive job without fatigue. When a display is mounted on the testing station, the driving circuits simulate a cellular phone to generate designed patterns on the display; image acquisition devices working at high speed capture these patterns and transfer the images to a software system to analyze.

Figure 1 shows images of two cellular phone displays. That on the left is an eligible product, compared with a defective product on the right. This display usually contains nearly 10 000 pixels and each can be lit individually. It is not common for all pixels to malfunction, but patterns such as missing pixels, bad pixel clusters, dark corner, etc., occur fairly often. These failures are all embodied by some local features that influence the uniformity of luminance intensity. If we take the average of the luminance intensity of all pixels, the local failure features will be averaged out by a huge sample size. This would lead to an inefficient reflection of such failures.

In this application, each display is independent of the others; each pixel in a display can be driven by the circuits individually and is also independent of other pixels. Under the in-control status, the luminance of each pixel follows a normal distribution with identical but unknown parameters. One display is thus a sample with 10 000 independent observations. However, the aforementioned defects make changes on the distribution of

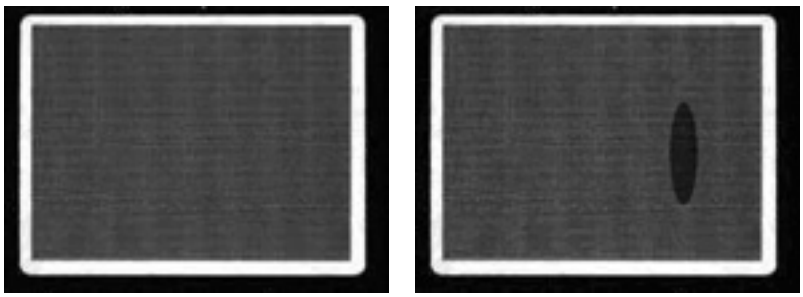

Figure 1. Sample images of the eligible and defective displays

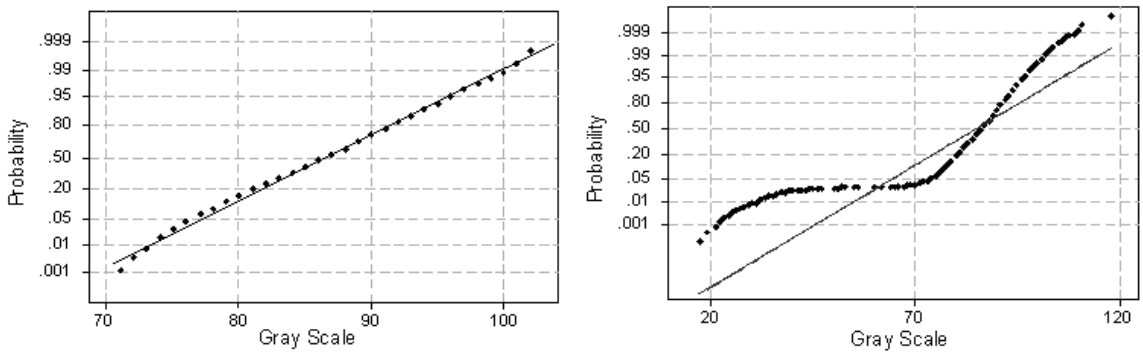

Figure 2. Q–Q plot of regular and defective displays

partial pixels and violate the normality of the sample. A Q–Q plot (i.e. a normal plot) of the sample is expected to observe these failures.

Figure 2 shows the Q–Q plots corresponding to the two displays shown in Figure 1. The differences in the plots are obvious. In a latter part of this paper, a straight line fitted to the Q–Q plot will be monitored to alarm the system to such failures.

1.2. Review of profile monitoring techniques

In many practical situations, the quality of a process or product is better characterized and summarized by the relationship between a response variable and one or more explanatory variables; that is, a profile. It is assumed that for each profile n ($n > 1$) values of the response variable (Y) are measured along with the corresponding values of one or more explanatory variables (X), which can be the time of operation, distance along a dimension of an object, or a general variable to related to the profile (see Woodall *et al.*[2]). Thus, profile monitoring tries to watch the relationship of the response variable and explanatory variables and draw conclusions about the quality characteristics.

Profiles are both widely observed in real applications and studied in a broad area of the literature. For example, a dissolution profile is a widely studied smoothing increasing curve in the pharmaceutical industry. The explanatory variable is the time elapsed from the beginning of the experiment and the response is the percentage of drugs dissolved up until the measuring epoch. Freitag[3] introduced several empirical indices for production line evaluation; most are developed based on experience rather than theoretical induction.

Walker and Wright[4] measured the quality of a particleboard along its depth direction, which forms a convex profile. The explanatory variable is the distance along the depth direction of the board and the response is the corresponding density. An additive model was proposed to compare profiles from different boards and different batches of boards. Conclusions can be drawn based on analysis of variance (ANOVA) analysis. The proposed method is potentially useful for profile monitoring.

A current–voltage (C–V) curve, which is formed by measuring the response variable current when applying explanatory variable voltage on a capacitor, has been used as a fast and effective method in micro-electric fabrication for quality control purposes for decades (see Walls *et al.*[5] and Lin *et al.*[6]). Cai and Sah[7] also introduced an approach to monitor process-residue interface traps of metal-oxide semiconductor (MOS) transistors after fabrication. By changing the gate voltage, the body current can be measured as an indication of the interface traps, which again forms a profile.

A linear profile is a first-order function to characterize the relationship between a response variable and explanatory variables. It is also the most widely observed class of profiles. Calibration processes are often characterized by linear functions (Woodall *et al.*[2]). Most processes with aging effect are estimated or approximated using a linear function for run-to-run control (see Buter and Stefani[8]). Many nonlinear relationships can be transformed into linear relationships by proper variable transformation and substitution (see Hastie *et al.*[9]).

However, there has been little research effort focused on how to efficiently monitor profiles. Several works of research have started from the linear profile monitoring issues. Kang and Albin[10] investigated a semiconductor manufacturing process and monitored a simple linear profile. Since the data that fit the linear profile can be characterized by a intercept, slope and residual, they proposed a method to treat the intercept and slope as two correlated random variables and convert the situation to a general bivariate problem, Hotelling's T^2 chart is then used for this purpose. They also considered the fitted residuals as a measurement of dispersion of the data from the line, and an exponentially weighted moving average residuals (EWMA-R) chart is used to monitor the residuals.

Kim *et al.*[11] transformed the original intercept and slope variables proposed by Kang and Albin[10] to release the correlation. Three EWMA charts are used to monitor the new uncorrelated intercept, slope and residual simultaneously. More complex profiles are also being studied. Woodall *et al.*[2] gave a broad review of the current research branches and further topics that need more attention.

As motivated by the cellular phone display failures that are distinguished by the linear profile in the Q–Q plot, we adopt and modify the existing linear profile monitoring techniques to fit the data-rich problem.

2. SPC USING THE Q–Q PLOT

The Q–Q plot is a graphical technique for determining whether two data sets come from populations with a common distribution (see NIST/SEMATECH[12]). It is generated by plotting the quantiles of the first data set against the quantiles of the second data set.

The Q–Q plot has long been a popular graphical technique for visually checking the distribution of a dataset or comparing two datasets (see Chambers *et al.*[13] and D'Agostino and Stephens[14]). Its power, however, has been underemployed. The application of Kittlitz and Sullivan[15] is a successful example in using a normal plot to draw important findings.

If the second data set is fixed as reference, the Q–Q plot can compare the distributional differences between many data sets by one-by-one comparison with the reference. If the reference data set and the compared data set follow the same distribution, a linear trend will be observed; if not, versatile modality may be shown, depending on the difference between the data sets.

The normal plot is a special case of the Q–Q plot, which assumes normal distribution of both data sets. As in the example shown above, failures occurring in the display, such as abnormal dots or clusters, dark corners or missing pixels, do violate the normality assumption of the sample and can be observed by the Q–Q plot. These failure patterns that lead to partial mean shifts or variance changes, although not easily distinguished from the sample mean, are expected to be alarmed by investigating and monitoring the Q–Q plot.

Many distributional aspects can be simultaneously tested by a Q–Q plot. For different distributions, the fitted line to the Q–Q plot conveys different meanings. As tabulated by Chambers *et al.*[13], for χ^2 distributed data, the intercept is always zero, while the slope estimates the cubic root of the distribution parameter. For a normal distribution, the intercept and slope can be interpreted as the mean and standard deviation of the sample data. Figure 3 illustrates the relationship of three groups of data generated from normal distributions with different

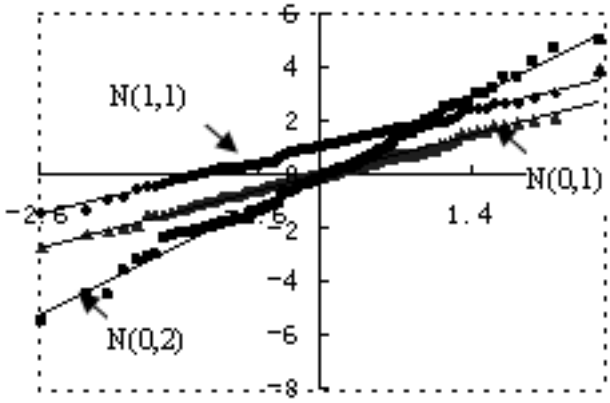

Figure 3. Q–Q plot of normal distribution with different parameters

means and standard deviations. The Q–Q plots and the fitted lines are drawn on the same graph. As can be seen, compared with the fitted line from a Normal(0, 1) dataset, the line fitted for the dataset from a Normal(1, 1) distribution shifts upward by one unit, while the line from the Normal(0, 2) dataset is steeper than the other two and it still passes through the origin.

A linear profile can be characterized by its intercept and slope. Parameter estimations are done by using generalized least square. Although the original observations are independent and have an identical distribution, the independence no longer holds when they are reordered and put in pairs on the Q–Q plot, nor are they of the same distribution. Generalized least squares can be applied to such a situation to estimate desired parameters. Stuart and Ord[16,17] gave a systematic introduction to ordered statistics and generalized least squares. Let $y_1, y_2, \ldots, y_n$ be n observations from the population distributed as normal with mean μ and standard deviation σ. Reorder the data from the smallest to the largest and denote each entry as $y_{(1)}, y_{(2)}, \ldots, y_{(n)}$, where $y_{(r)}$ is the rth smallest value. Let $z_{(r)}$ be the corresponding quantile from the standard normal distribution. There exists a relationship between $y_{(r)}$ and $z_{(r)}$ (see Stuart and Ord[16,17]):

$$y_{(r)} = \mu + \sigma z_{(r)}, \quad r = 1, 2, \ldots, n \tag{1}$$

For sufficiently large sample size n, the vector $\mathbf{z} = (z_{(1)}, z_{(2)}, \ldots, z_{(n)})$ is approximately multivariate normally distributed. Let $E(\mathbf{z}) = \mathbf{m} = (m_1, m_2, \ldots, m_n)'$ and $\mathrm{Cov}(\mathbf{z}) = \mathbf{V}$. The calculations for the elements of the variance–covariance matrix of $\mathbf{z}$ are

$$v_{ij} = \frac{p_i q_j}{n f_i f_j}, \quad \text{for all } 1 \le i \le j \le n \tag{2}$$

where $p_i = \Phi(z_{(i)})$ is the percentage of the data less than $z_{(i)}$ for the standard normal distribution, $q_i = 1 - p_i$, $f_i \equiv f(z_{(i)}) = \mathrm{d}\Phi(z_{(i)})/\mathrm{d}x$ is the density function of the standard normal distribution at point $z_{(i)}$. The above relationship defines the upper triangle of the matrix; the lower half can be obtained from the symmetric property by transposing.

Correspondingly, $\mathbf{y} = (y_{(1)}, y_{(2)}, \ldots, y_{(n)})'$ has $E(\mathbf{y}) = \mu\mathbf{1} + \sigma\mathbf{m}$ and $\mathrm{Cov}(\mathbf{y}) = \sigma^2\mathbf{V}$, where $\mathbf{1}$ is the unit vector. It is easy to see that $y_{(r)}$ has a different mean and variance for different r. $\mathbf{y}$ can be re-expressed as

$$\mathbf{y} = \mu\mathbf{1} + \sigma\mathbf{m} + \varepsilon \tag{3}$$

where ε is the error term and $E(\varepsilon) = \mathbf{0}$ and $E(\varepsilon\varepsilon') = \sigma^2\mathbf{V}$.

Under the assumption of multivariate normality of $\mathbf{y}$, we can use generalized least squares to fit the regression model and find the parameters μ and σ. For a group of observations $\mathbf{y}$ that are sampled from a normal

distribution, the estimation for the intercept and slope parameters of the fitted line are (Stuart and Ord[16,17])

$$\hat{\mu} = \mathbf{1}'\mathbf{V}^{-1}\mathbf{y}/\mathbf{1}'\mathbf{V}^{-1}\mathbf{1} \tag{4}$$

$$\hat{\sigma} = \mathbf{m}'\mathbf{V}^{-1}\mathbf{y}/\mathbf{m}'\mathbf{V}^{-1}\mathbf{m} \tag{5}$$

$$\mathrm{Var}(\hat{\mu}) = \sigma^2/\mathbf{1}'\mathbf{V}^{-1}\mathbf{1} \tag{6}$$

$$\mathrm{Var}(\hat{\sigma}) = \sigma^2/\mathbf{m}'\mathbf{V}^{-1}\mathbf{m} \tag{7}$$

$$\mathrm{Cov}(\hat{\mu}, \hat{\sigma}) = 0 \tag{8}$$

As can be seen from (4)–(8), the covariance between the intercept and the slope is zero, which enables us to simplify the monitoring by using an univariate chart later on. If given V^{-1}, for a group of observations, the estimation for $\hat{\mu}$ and $\hat{\sigma}$ can easily be obtained.

By investigating the residual, we are able to estimate how well the line fits the data. Because of the unequal variance and correlation between the ordered statistics, we use the error sum of square (ESS) as an indication:

$$\mathrm{ESS} = (\mathbf{y} - \hat{\mathbf{y}})'\mathbf{V}^{-1}(\mathbf{y} - \hat{\mathbf{y}}) \tag{9}$$

where $\hat{\mathbf{y}}$ is the estimation of the observations and satisfies the following relationship:

$$\frac{(\mathbf{y} - \hat{\mathbf{y}})'\mathbf{V}^{-1}(\mathbf{y} - \hat{\mathbf{y}})}{\sigma^2} \sim \chi^2_{n-2} \tag{10}$$

The ESS measures how well the line fits the data and judges the linearity of the samples.

3. PHASE I METHODS

Woodall[18] has emphasized distinguishing between Phase I and Phase II when setting up a control chart. Phase I is a procedure that helps us to understand the variation of a process and model the in-control process after stabilizing it. A group of historical observations are usually given in Phase I. We need to determine the status of the process and keep only those observations that are taken when the process is in-control. Then, process parameters such as the mean and variance are estimated from the in-control historical observations.

One important job in Phase I is to separate in-control and out-of-control data from historical observations. In our application, we use separate Shewhart-type $\bar{x}$–R charts for the intercept, slope and ESS. Observations that exceed the control limit are removed and then the control limits are adjusted to check the data set again. This step is iterated several times until all points are in-control. Once the stabilization is achieved, parameters will be estimated by proper modeling.

For the large sample size scenario this paper targets, the linear profile generated from the Q–Q plot is used to model a stable process in both Phase I and Phase II.

Parameters of the model include its intercept μ and slope σ. We use this notation because the intercept and slope of the fitted line can be interpreted as the mean and standard deviation of the sample (see Chambers *et al.*[13]).

It is necessary to estimate the target value of μ, σ and ESS from the historical data. Suppose that k groups of in-control samples are left after the above stabilization step. For each sample, a line is fitted to its Q–Q plot and a group of μ, σ and ESS are estimated. After obtaining k groups of values, the average of these is taken as the target of the parameters. Denote μ_0 and σ_0 as the target value of μ and σ, respectively, and the following relationships can easily be obtained (see Kang and Albin[10]),

$$\mu_0 = \frac{\sum_{i=1}^{k} \mu_i}{k}, \quad \sigma_0 = \frac{\sum_{i=1}^{k} \sigma_i}{k} \tag{11}$$

We then can replace the σ in (6) and (7) by σ_0 for use in Phase II.

As suggested by Kim *et al.*[11], the approach proposed by Crowder and Hamilton[19] to monitor residuals works efficiently. Under the assumption of multivariate normality of $\mathbf{y}$, $\mathrm{ESS} = (\mathbf{y} - \hat{\mathbf{y}})'\mathbf{V}^{-1}(\mathbf{y} - \hat{\mathbf{y}})$ follows approximately a χ^2 distribution with $n - 2$ degrees of freedom. As proved by Crowder and Hamilton[19], if the logarithmic transformation of ESS (denoted as LESS) is be charted, a better result is expected. Thus, for each sample we take $\mathrm{LESS}_i = \log(\mathrm{ESS}_i)$ as observations for charting. The mean and variance of LESS can also be estimated from the k groups of in-control samples:

$$\mathrm{LESS}_0 = \frac{\sum_{i=1}^{k} \mathrm{LESS}_i}{k}, \quad \mathrm{Var}(\mathrm{LESS}_0) = \frac{\sum_{i=1}^{k} (\mathrm{LESS}_i - \mathrm{LESS}_0)^2}{k-1} \tag{12}$$

Until now, we have transformed the large sample into a linear profile and obtained an estimation for the parameters. Profile monitoring techniques are employed in the next phase to monitor the fitted line and alarm failures of the samples.

4. PHASE II METHODS

Given the target value of parameters, the task in Phase II is to set up charts to monitor the variation of the parameters and detect hidden shifts based on the target condition obtained from Phase I.

The fitted line to the Q–Q plot from one sample can be characterized by its intercept and slope. As can be learned from (8), the estimators of the intercept and slope are independent; this feature benefits us in that it simplifies the monitoring. Two individual EWMA charts will be used to monitor the intercept and slope simultaneously. We denote this monitoring scheme as the 2–E method.

As the normality of samples is better seen in the linearity of the Q–Q plot, we also add a chart to monitor the residual. This happens to be what Kim *et al.*[11] did. In their method, three EWMA charts are combined to monitor the intercept, slope and residual. One minor modification to fit our application is the estimation of residual. The residual is estimated by the ESS, as shown in (9). We denote this monitoring scheme as the 3–E method.

The only difference between the 2–E and 3–E methods is whether or not the residual is monitored. Adding one chart to monitor the residual may increase the false alarm rate. Manually decreasing the false alarm rate by restricting the in-control average run length (ARL) may, however, lower the detection power. The tradeoff will be judged in the next section by Monte Carlo simulation.

To set up EWMA charts for intercept, slope and residual, we use the following recursive relationships and control limits:

$$\begin{cases} \mathrm{EWMA}_\mu(i) = r\mu_i + (1-r)\mathrm{EWMA}_\mu(i-1) \\ \mathrm{CL}_\mu = \mu_0 \pm k_\mu \sigma_\mu \sqrt{\dfrac{r}{2-r}} \end{cases} \tag{13}$$

$$\begin{cases} \mathrm{EWMA}_\sigma(i) = r\sigma_i + (1-r)\mathrm{EWMA}_\sigma(i-1) \\ \mathrm{CL}_\sigma = \sigma_0 \pm k_\sigma \sigma_\sigma \sqrt{\dfrac{r}{2-r}} \end{cases} \tag{14}$$

$$\begin{cases} \mathrm{EWMA}_e(i) = r\,\ln(\mathrm{ESS}_i) + (1-r)\mathrm{EWMA}_e(i-1) \\ \mathrm{CL}_e = \mathrm{LESS}_0 \pm k_e \sigma_e \sqrt{\dfrac{r}{2-r}} \end{cases} \tag{15}$$

for $i = 1, 2, \ldots$. μ_0, σ_0 and LESS_0 are the target of intercept, slope and the logarithmic ESS, respectively. σ_μ, σ_σ and σ_e are the standard deviations corresponding to each parameter. r is the smoothing parameter for EWMA chart. k_μ, k_σ and k_e are set to achieve the desired in-control ARL for each chart. The starting value of each chart is set as the corresponding target value μ_0, σ_0 and LESS_0, respectively.

Two-sided EWMA charts are used since we want to be alarmed whenever the process deviates from the normal to either a better or worse condition.

5. RUN LENGTH PERFORMANCE ANALYSIS

In order to investigate monitoring performance, we use Monte Carlo simulation to find the influence of mean shifts and standard deviation shifts with different patterns.

Three methods are selected for comparison. The first is the aforementioned 2–E method. Two univariate EWMA charts are employed to monitor the intercept and slope. The second is the 3–E method, with an additional chart to simultaneously monitor the residual. The weight coefficient in all EWMA charts is set to 0.2, which is a moderate value adopted by most studies in the literature. Finally, the conventional $\bar{x}$–S chart is also set for the same group of data.

Without loss of generality, all observations are generated from the normal distribution with $\mu = 0$ and $\sigma = 1$. The size of each sample is 500. All ARL values are estimated using at least 10 000 replicates.

The overall in-control ARLs of all the combined charts are set to 200. However, for the 3–E method, the individual in-control ARL of each EWMA chart is roughly 586. For the other two methods, each of the individual charts has an in-control ARL of about 400.

As mentioned above, a common failure condition of the display corresponds to a partial mean shift or partial standard deviation shift, so we investigate the performance of different charting methods by adjusting the shift proportion from 100% to a smaller value.

Table I gives the ARLs for shifts in sample mean. The third column lists ARLs as corresponding to a 100% mean shift. The left columns show the results corresponding to partial mean shifts; that is, less than 100% of observations in one sample have shifted, while others remain unchanged. For the 100% mean shift case, compared with the two methods that use the EWMA chart, the $\bar{x}$–S chart is slower in detecting the mean shift when shift size is smaller than 2σ. However, if shift size become larger, $\bar{x}$–S becomes as good as or even better than the other two EWMA-based methods. When the shift proportion decreases from 100%, the $\bar{x}$–S method is no longer competitive with the other two methods, even when the shift size is large.

Among the two EWMA-based methods, the 2–E method is superior to or competitive with the 3–E method in all cases. This result shows that when adding one more EWMA chart, the effect of increasing the detection power is not as strong as of increasing the false alarm rate. Forcing the in-control ARL to be equal, the detecting power of the 3–E method becomes lower than that of the 2–E method.

Table II shows the ARLs of the three methods when standard deviation shifts occur. Based on the valuable suggestion of anonymous reviewers, a single EWMA chart to monitor the ESS was set up. This experiment reveals that a single ESS chart does not show obvious improvement in monitoring the mean shifts, but it does have the capability to alarm standard deviation shifts. So in Table II, we add the results from a single EWMA chart that monitors the ESS for comparison.

The column title still indicates the shift proportion, decreasing from 100% to 20%. $\bar{x}$–S methods still perform worse than the other two and a little better when shift size becomes larger than about 1.20, and this superiority vanishes gradually when the shift proportion decreases.

The ESS chart shows capability to alarm standard deviation shifts for all proportions. However, when combining with the 2–E method to generate a 3–E method, the advantage is not so obvious. The 3–E method is competitive with, but not always better than, the 2–E method. This shows that the additional chart within the 3–E method does not contribute much more to the detection of shift than the false alarm it introduced.

6. CONCLUSION

This work has aimed to monitor a huge sample sized scenario. Novel schemes based on Q–Q plot transformation have been proposed to deal with some serious failure patterns.

Table I. ARL comparison of three schemes for mean shift from $\mu_{\bar{y}} = 0$ to $\mu_{\bar{y}} = \alpha$

α	Charting methods	Shift proportion				
		100%	80%	60%	40%	20%
0.20	$\bar{x}$–S	180.0	185.6	191.9	195.6	199.6
	3–E	131.2	150.0	170.3	185.2	196.2
	2–E	120.1	141.1	163.8	182.6	196.5
0.40	$\bar{x}$–S	139.4	157.0	173.1	185.7	195.4
	3–E	57.7	80.4	112.5	151.4	185.5
	2–E	50.3	70.9	101.7	141.0	182.6
0.60	$\bar{x}$–S	96.7	121.0	147.9	173.2	191.8
	3–E	27.8	42.2	68.0	112.5	170.0
	2–E	24.5	37.0	59.5	101.8	164.1
0.80	$\bar{x}$–S	63.7	89.4	121.2	157.4	185.7
	3–E	16.1	24.6	42.1	80.5	151.7
	2–E	14.4	21.9	37.0	71.0	141.3
1.00	$\bar{x}$–S	40.9	63.8	96.8	139.6	180.1
	3–E	10.6	16.1	27.7	57.8	132.0
	2–E	9.7	14.5	24.6	50.4	120.5
1.50	$\bar{x}$–S	15.2	27.4	50.9	96.8	161.4
	3–E	5.6	7.8	12.8	27.6	88.1
	2–E	5.3	7.3	11.7	24.6	77.6
2.00	$\bar{x}$–S	6.5	12.6	27.4	63.6	138.9
	3–E	3.8	5.1	7.8	16.1	57.7
	2–E	3.6	4.8	7.3	14.5	50.5
2.50	$\bar{x}$–S	3.3	6.5	15.2	40.8	116.6
	3–E	2.9	3.8	5.6	10.6	39.1
	2–E	2.8	3.6	5.3	9.7	34.4
3.00	$\bar{x}$–S	2.0	3.7	8.9	27.4	96.1
	3–E	2.4	3.1	4.4	7.8	27.8
	2–E	2.3	2.9	4.1	7.3	24.6

Three approaches, the $\bar{x}$–S, 2–E and 3–E methods, have been applied to monitor both mean and standard deviation shifts. A single ESS chart has also been set up for the standard deviation shifts. The tradeoff of adding one chart to increase the false alarm rate and increase the detecting power has also been investigated using simulation.

Simulations were conducted for the cases where different shift proportions are applied for the main findings are summarized below.

1. For a mean shift, the EWMA-based method is superior to the $\bar{x}$–S method when the shift size is small, while the superiority gradually vanishes when the shift proportion decreases from 100%, which is usually the case for a large sample size scenario. The 2–E method has always outperformed the 3–E method, suggesting that adding one more chart to monitor the residual does not contribute much to the detection of the mean shift, but leads to an increase in false alarm rate. Thus, the 2–E method is recommended for the mean shift monitoring.
2. The $\bar{x}$–S method is slow in detecting the standard deviation shift, while a single EWMA chart monitoring the ESS is better than the $\bar{x}$–S method. The 2–E method shows the best performance in almost all cases. Although the 3–E is competing with the 2–E method, especially for a small shifts with small proportion, adding one more chart will also increase the complexity. Thus, the recommendation is still to use the 2–E method for a general situation. A single ESS chart can be considered for the sake of simplicity of application.

Table II. ARL comparison of three schemes for standard deviation shift from $\sigma_y = 1$ to $\sigma_y = \beta$

β	Charting methods	Shift proportion 100%	80%	60%	40%	20%
1.02	$\bar{x}$–S	76.7	99.1	127.0	157.3	183.8
	3–E	27.7	42.8	69.8	116.5	177.3
	2–E	26.7	41.5	69.9	119.4	184.0
	ESS	46.1	63.5	90.2	130.9	178.0
1.04	$\bar{x}$–S	20.5	34.1	57.8	97.6	156.0
	3–E	7.7	11.2	19.1	41.7	115.0
	2–E	7.4	10.7	18.3	40.9	118.2
	ESS	14.5	20.9	33.9	62.4	129.0
1.06	$\bar{x}$–S	6.9	12.5	25.5	56.5	124.7
	3–E	4.3	5.8	9.0	18.8	67.1
	2–E	4.1	5.6	8.6	18.0	67.4
	ESS	7.5	10.5	16.8	33.5	87.9
1.08	$\bar{x}$–S	3.1	5.7	12.2	32.3	95.4
	3–E	3.0	3.9	5.7	10.8	40.4
	2–E	2.9	3.7	5.5	10.4	39.0
	ESS	5.1	6.8	10.3	20.4	60.4
1.10	$\bar{x}$–S	1.8	3.0	6.5	18.9	72.0
	3–E	2.4	3.0	4.2	7.4	25.7
	2–E	2.3	2.9	4.0	7.1	24.7
	ESS	3.8	5.0	7.3	13.7	43.1
1.20	$\bar{x}$–S	1.0	1.0	1.3	2.7	16.4
	3–E	1.2	1.5	2.0	2.8	6.9
	2–E	1.1	1.4	1.9	2.7	6.7
	ESS	1.9	2.3	3.0	4.7	12.7
1.30	$\bar{x}$–S	1.0	1.0	1.0	1.2	5.0
	3–E	1.0	1.0	1.2	1.9	3.8
	2–E	1.0	1.0	1.2	1.8	3.6
	ESS	1.3	1.6	2.0	2.9	6.4
1.40	$\bar{x}$–S	1.0	1.0	1.0	1.0	2.2
	3–E	1.0	1.0	1.0	1.4	2.6
	2–E	1.0	1.0	1.0	1.3	2.5
	ESS	1.0	1.2	1.6	2.1	4.2
1.50	$\bar{x}$–S	1.0	1.0	1.0	1.0	1.4
	3–E	1.0	1.0	1.0	1.1	2.1
	2–E	1.0	1.0	1.0	1.1	2.0
	ESS	1.0	1.0	1.2	1.7	3.1

3. The monitoring performances of all charting schemes decrease rapidly when the shifted proportion decreases because of the submerging effect of the large sample size.

A notable promising usage of the Q–Q plot is to monitor a process whose output is not normal. The traditional efforts in monitoring non-normal outputs are either to transform the original data to a normal distribution or to use non-symmetrical control limits to fit non-normality (see Yourstone and Zimmer[20] and Chou *et al.*[21]). By using the Q–Q plot monitoring method, we can easily handle the non-normality of the process by changing the reference distribution or simply use empirical distribution if no existing distributions can be fitted to the data perfectly. Further research could extend the proposed scheme to multivariate scenarios using similar techniques to those presented in this paper.

Profile monitoring is a promising area, which has a wide application in reality. Research on the statistical monitoring of profiles, however, has just begun. The linear profile monitoring techniques provide efficient ways to monitor the huge sample size environment. We believe that research on the statistical monitoring of more general profiles is extreme challenging and will benefit industrial applications greatly.

Acknowledgements

We thank the editor and the referees for their valuable comments that have improved this work. This research was supported by RGC Competitive Earmarked Research Grants HKUST6183/03E and HKUST6232/04E.

REFERENCES

1. Barry MJA, Linoff GS. *Mastering Data Mining: The Art and Science of Customer Relationship Management*. Wiley: New York, 2000.
2. Woodall WH, Spitzner DJ, Montgomery DC, Gupta S. Using control charts to monitor process and product quality profiles. *Journal of Quality Technology* 2004; **36**:309–320.
3. Freitag G. Guidelines on dissolution profile comparison. *Drug Information Journal* 2001; **35**:865–874.
4. Walker E, Wright SP. Comparing curves using additive models. *Journal of Quality Technology* 2002; **34**:118–129.
5. Walls JA, Walton A, Roertson JM, Crawford TM. Automating and sequencing C–V measurements for process fault diagnosis using a pattern-recognition approach. *Proceedings of the 1989 International Conference on Microelectronic Test Structures*, Edinburgh, U.K., March 1989. IEEE Press: New York, 1989; 193–199.
6. Lin H, Chien C, Chen C, Hsien S, Wang M, Huang T, Chang C. Evaluation of plasma charging damage in ultrathin gate oxides. *IEEE Electron Device Letters* 1998; **19**:68–70.
7. Cai J, Sah C. Monitoring interface traps by DCIV method. *IEEE Electron Device Letters* 1999; **20**:60–63.
8. Butler SW, Stefani JA. Supervisory run-to-run control of polysilicon gate etch using *in situ* ellipsometry. *IEEE Transactions on Semiconductor Manufacturing* 1994; **7**:193–201.
9. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer: New York, 2001.
10. Kang L, Albin SL. On-line monitoring when the process yields a linear profile. *Journal of Quality Technology* 2000; **32**:418–426.
11. Kim K, Mahmoud MA, Woodall WH. On the monitoring of linear profiles. *Journal of Quality Technology* 2003; **35**:317–328.
12. NIST/SEMATECH. e-Handbook of Statistical Methods. http://www.itl.nist.gov/div898/handbook/ [November 2003].
13. Chambers J, Cleverland WS, Kleiner B, Tukey PA. *Graphical Methods for Data Analysis*. Chapman and Hall/CRC: Boca Raton, FL, 1998.
14. D'Agostino RB, Stephens MA. *Goodness-of-fit Techniques*. Marcel Dekker: New York, 1996.
15. Kittlitz R, Sullivan J. Letter to the editor—detection of multiple change points from clustering individual observations. *Journal of Quality Technology* 2003; **35**:237–238.
16. Stuart A, Ord J. *Kendall's Advanced Theory of Statistics*, vol. 1. Oxford University Press: New York, 1991.
17. Stuart A, Ord J. *Kendall's Advanced Theory of Statistics*, vol. 2. Oxford University Press: New York, 1991.
18. Woodall W. Controversies and contradictions in statistical process control. *Journal of Quality Technology* 2000; **32**:341–378.
19. Crowder SV, Hamilton MD. An EWMA for monitoring a process standard deviation. *Journal of Quality Technology* 1992; **24**:12–21.
20. Yourstone SA, Zimmer WJ. Non-normality and the design of control charts for averages. *Decision Sciences* 1992; **23**:1099–1113.
21. Chou Y, *et al*. Transforming non-normal data to normality in statistical process control. *Journal of Quality Technology* 1998; **30**:133–141.

Authors' biographies

Kaibo Wang is a PhD candidate in the Department of Industrial Engineering and Engineering Management, Hong Kong University of Science and Technology. He received his BS and MS degrees from Xi'an Jiaotong University, Xi'an, Peoples Republic of China, in Mechatronics Engineering. His research project focuses on the statistical monitoring and control of industrial processes with mixed-resolution information.

Fugee Tsung is an Associate Professor of Industrial Engineering and Engineering Management at HKUST. He received both his MS and PhD in Industrial and Operations Engineering from the University of Michigan, Ann Arbor, and his BSc in Mechanical Engineering from the National Taiwan University. He worked for the Ford Motor and Rockwell International and carried out his post-doctoral research with Chrysler. He is the Chair of the Quality, Statistics, and Reliability (QSR) Section of the Institute for Operations Research and the Management Sciences (INFORMS), and is currently on the editorial boards for the *International Journal of Reliability, Quality and Safety Engineering (IJRQSE)* and the *International Journal of Six Sigma and Competitive Advantage (IJSSCA)*. He was the winner of the Best Paper Award for the *IIE Transactions Focus Issues on Quality and Reliability* in 2003, and is also an ASQ Certified Six Sigma Black Belt and ASQ Authorized Six Sigma Master Black Belt Trainer. His research interests include quality engineering and management, process control and monitoring, and Six Sigma implementation.