18. Multivariate Statistical Process Control Schemes for Controlling a Mean

The quality of products produced and services provided can only be improved by examining the process to identify causes of variation. Modern production processes can involve tens to hundreds of variables, and multivariate procedures play an essential role when evaluating their stability and the amount of variation produced by common causes. Our treatment emphasizes the detection of a change in level of a multivariate process.

After a brief introduction, in Sect. 18.1 we review several of the important univariate procedures for detecting a change in level among a sequence of independent random variables. These include Shewhart's X–bar chart, Page's cumulative sum, Crosier's cumulative sum, and exponentially weighted moving-average schemes.

Multivariate schemes are examined in Sect. 18.2. In particular, we consider the multivariate T^2 chart and the related bivariate ellipse format chart, the cumulative sum of T chart, Crosier's multivariate scheme, and multivariate exponentially weighted moving-average schemes.

An application to a sheet metal assembly process is discussed in Sect. 18.3 and the various multivariate procedures are illustrated.

Comparisons are made between the various multivariate quality monitoring schemes in Sect. 18.4. A small simulation study compares average run lengths of the different procedures under some selected persistent shifts.

When the number of variables is large, it is often useful to base the monitoring procedures on principal components. Section 18.5 discussesthis approach. An example is also given using the sheet metal assembly data.

Finally, in Sect. 18.6, we warn against using the standard monitoring procedures without first checking for independence among the observations. Some calculations, involving first-order autoregressive dependence, demonstrate that dependence causes a substantial deviation from the nominal average run length.

Today, with automated data collection a common practice, data on many characteristics need to be continually monitored. In this chapter, we briefly review the major univariate methods and then discuss the multivariate quality monitoring methods for detecting a change in the level of a process. We concentrate on the sequential schemes where the average run length curve is of primary importance for describing the performance. All univariate monitoring schemes attempt to determine if a sequence of observations $X_1, X_2, \cdots$ is stable. That is, to confirm that the mean and variance remain constant. Throughout our discussion we assume that the X_i are in-

dependent. However, in the last section, we do consider the effect of dependence on average run length.

We begin with a review of univariate procedures in Sect. 18.1 and then go on to multivariate extensions in Sect. 18.2. In Sect. 18.3, an example is used to illustrate the behavior of different multivariate procedures. The performance of the multivariate schemes is also compared via simulation in Sect. 18.4. Control charts based on principle components are introduced in Sect. 18.5. In Sect. 18.6, we discuss the difficulties caused by time dependence in the sequence of observations.

18.1 Univariate Quality Monitoring Schemes

To set notation, let $X_1, X_2, \cdots$ be the sequence of independent random variables produced by a process being monitored. Let a denote the target mean value. In the fixed-sample-size setting, the hypothesis to be tested is $H_0 : \mu = a$ versus $H_1 : \mu \neq a$. However, in the sequential setting, we develop statistics based on the deviation from the target value. The statistics will all involve s, an estimate of the standard deviation of a single observation.

Typically, the sequential schemes involve a constant k, called a reference value, and a positive constant h, that defines the decision rule. The two constants (k, h) are selected to give good average run length (ARL) properties. Sometimes, separate reference values k^+ and k^- are used to detect an increase and a decrease in the mean, respectively.

Figure 18.1 shows 50 observations collected on an automotive sheet metal assembly process. These measurements, made by sensors, are deviations from the nominal values (millimeters) at the back right-hand side of the car body. Measurements made at various locations of the car body are presented along with more details in Sect. 18.4.

Figures 18.2–18.5 present different univariate statistics for detecting a change within that data.

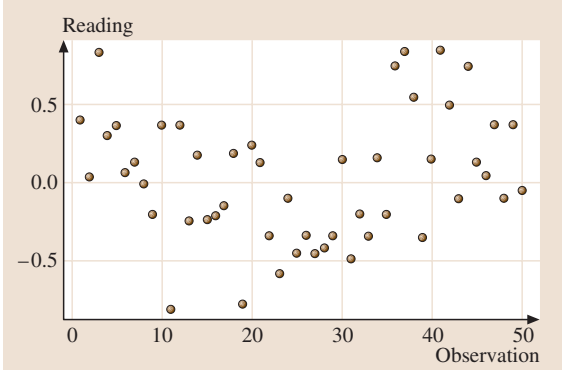

Fig. 18.1 Automotive assembly data

18.1.1 Shewhart *X*-Bar Chart

Historically, the first quality control chart was defined by *Shewhart* [18.1]. This procedure has been widely used since the 1940s. The Shewhart X-bar chart is generally based on the mean of a small sample. That is, the plotted point is usually the mean $\overline{x}_n$ of a few $(m \geq 1)$ observations. When samples are available, rather than just individual observations, charts are also maintained to monitor the process standard deviation.

A widely used Shewhart's X-bar chart [18.1] signals a shift in mean when

$$\overline{X}_n \geq a + 3\frac{s}{\sqrt{m}} \quad \text{or} \quad \overline{X}_n \leq a - 3\frac{s}{\sqrt{m}} ,$$

where s is an estimate of the standard deviation of an individual observation obtained from data collected during stable operation.

Figure 18.2 illustrates Shewhart's X-bar chart applied to automotive assembly data with $m = 1$.

Figure 18.1 suggests that there is a small increase in the mean towards the end of the sequence. However, it is not detected by the Shewhart's X-bar chart.

The Shewhart X-bar chart is very simple and effective for detecting an isolated large shift. However,

Fig. 18.2 Shewhart's X-bar chart $(m = 1)$ applied to automotive assembly data

because any decision using the Shewhart chart is based only on the most recent observation and has no memory, it is not effective in detecting small or moderate shifts, even if the shifts are persistent.

To help remedy the insensitivity to small shifts, practitioners often apply additional rules for signaling a change. These conditions to signal include:

i) nine points in a row on the same side of the centerline

ii) six points in a row that are decreasing or six that are increasing.

18.1.2 Page's Two-Sided CUSUM Scheme

The univariate two-sided cumulative sum (CUSUM) scheme proposed by *Page* [18.2] uses two cumulative sums: one to detect an increase in mean and another to detect a decrease. It is based on single observations X_n rather than means.

For $n > 1$, Page iteratively defines the two statistics

$$S_{H(n)} = \max(0,\, S_{H(n-1)} + X_n - k^+)\,, \tag{18.1}$$

$$S_{L(n)} = \min(0,\, S_{L(n-1)} + X_n - k^-) \tag{18.2}$$

with specified starting values $S_{H(0)} \geq 0$, $S_{L(0)} \leq 0$. Separate reference values k^+ and k^- are used in Page's CUSUM scheme to detect an increase in the mean and a decrease in the mean, respectively. Specifically, k^+ is selected to be larger than the target a so each term in the sum has a slightly negative expected value, and k^- is selected to be smaller than a. When the sequence of random variables X_1, $X_2, \cdots$ are normally distributed, typically we can set $k^+ = a + ks$ and $k^- = a - ks$, where k is one half of the specified shift in mean (expressed in standard deviations) that should be quickly detected by the scheme.

If the process remains in control at the target value, the CUSUM statistics defined in (18.1) and (18.2) should vary randomly but stay close to zero. When there is an increase in the process mean, a positive drift will develop in the CUSUM statistics S_H. Conversely, if there is a decrease in the mean, then a negative drift will develop in S_L. Therefore, an increasing trend in S_H or a decreasing trend in S_L is taken to indicate a shift in the process mean.

Page's CUSUM scheme signals a shift in mean when

$S_{H(n)} \geq hs$ (signals an increase), or

$S_{L(n)} \leq -hs$ (signals a decrease) .

The positive constant h is chosen to obtain a desired value of in-control ARL.

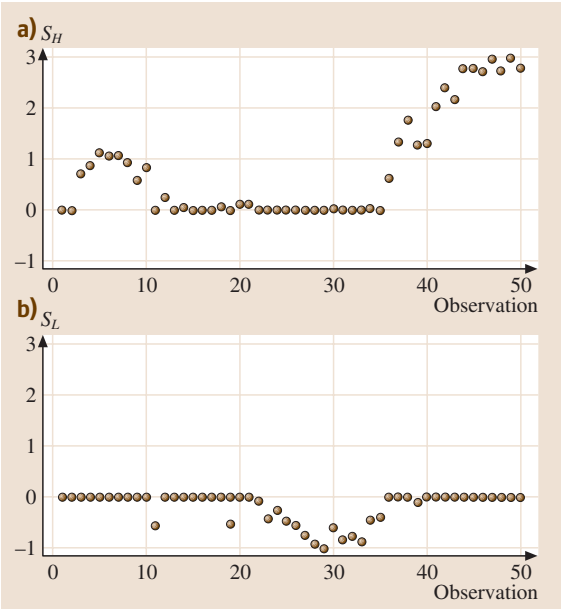

Fig. 18.3 Page's CUSUM statistics S_H (a) and S_L (b) using automotive assembly data

Figure 18.3 illustrates Page's CUSUM statistics applied to the automotive assembly data. The increasing trend in the plot of S_H indicates an increase in the mean towards the end of the sequence. Generally, Page's CUSUM scheme is more effective in detecting small but persistent shifts than the Shewhart X-bar chart.

18.1.3 Crosier's Two-Sided CUSUM Scheme

Crosier [18.3] proposed a two-sided CUSUM scheme which first updates the previous CUSUM by a new observation. Depending on the updated value of this sum, the new value of the CUSUM is either set equal to zero or the sum is shrunk towards zero. This modification reduces the chance of giving a false alarm. It seems from *Crosier* [18.3] that this procedure was arrived at empirically by trying a great many different schemes.

In particular, Crosier's two-sided CUSUM starts with $S_0 = 0$. For each step, first calculate the tentative sum $C_n = |S_{n-1} + (X_n - a)|$. Then Crosier's CUSUM statistic S_n is iteratively defined as

$$S_n = \begin{cases} 0 & \text{if } C_n \leq ks \\ (S_{n-1} + X_n - a)(1 - ks/C_n) & \text{otherwise.} \end{cases} \tag{18.3}$$

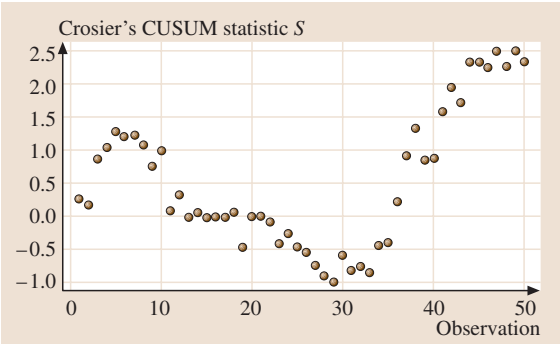

Fig. 18.4 Crosier's CUSUM statistic using automotive assembly data

Here, the constant k can also be set equal to one half of a specified mean-shift (expressed in standard deviations) that should be detected quickly.

Crosier's scheme signals that the mean has shifted when

$$S_n \geq hs \quad \text{(increase)}$$

or

$$S_n \leq -hs \quad \text{(decrease)} .$$

Figure 18.4 illustrates Crosier's two-sided CUSUM statistic applied to automotive assembly data.

The plot indicates a possible small decrease in the mean near the middle of the sequence and an increase towards the end.

18.1.4 EWMA Scheme

The univariate exponentially weighted moving-average (EWMA) scheme [18.4] is based on the weighted average of the current CUSUM and the new observation. The EWMA scheme smooths the sequence of observations by taking an average where the most recent observation receives the highest weight. Starting at $Z_0 = 0$, the updated EWMA sum is defined as

$$Z_n = r(X_n - a) + (1 - r)Z_{n-1} \qquad (18.4)$$

for $n = 1, 2, \cdots$, where $0 \leq r \leq 1$ is a specified constant. Expressed in terms of all the observations, we have

$$\begin{aligned} Z_n &= r\sum_{i=0}^{n-1}(1-r)^i(X_{n-i} - a) \\ &= r(X_n - a) + r(1-r)(X_{n-1} - a) \\ &\quad + r(1-r)^2(X_{n-2} - a) + \cdots , \end{aligned}$$

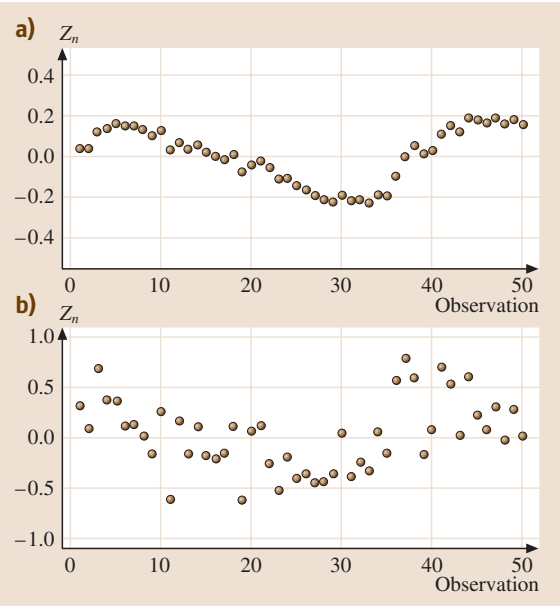

Fig. 18.5 EWMA with $r = 0.1$ (a) and $r = 0.8$ (b), using automotive assembly data

where the weights $r(1-r)^i$ fall off exponentially, giving rise to the name of the EWMA scheme.

The EWMA scheme signals a change in mean when

$$Z_n \geq hs \text{ (increase) or } Z_n \leq -hs \text{ (decrease)} .$$

Figure 18.5 shows the EWMA statistics with different choices of r applied to automotive assembly data. In the plot we can see that the EWMA statistic with the smaller value of parameter r is more sensitive to small shifts in this process mean. To be specific, the plot of the EWMA statistics with $r = 0.1$ has a decreasing trend in the middle and then an increasing towards the end of the sequence, which suggests a decrease and then an increase in the mean. However, in the plot of EWMA statistics with $r = 0.8$, such patterns can hardly be recognized.

As shown in Fig. 18.5, the choice of the value for r is critical to the performance of an EWMA scheme. Usually a small value for r is used for detecting a small shift. *Lucas* and *Saccucci* [18.4] extensively discuss the design of EWMA control schemes. For normal observations, they provide a table of *optimal* parameters r and h, which give the minimum ARL at the specified shift in process for specified in-control ARL. Lucas and Saccucci also compare the ARLs of the EWMA and Page's CUSUM schemes. They show that, with carefully chosen parameters r and h, the ARLs for EWMA are usually

smaller than the ARLs of Page's CUSUM scheme when the shift is smaller than the specified value for the shift that the scheme is designed to detect. However, the ARLs for the EWMA are larger than those of Page's CUSUM procedure when the shift in process mean is larger than the specified value, unless r is very large.

18.1.5 Summary Comments

Our review above included most of the popular univariate quality monitoring schemes. See [18.5] for further discussion of univariate schemes for process control. The well-known Shewhart X-bar chart is very effective for detecting large shifts but may be slow to detect small or moderate shifts. This fact gave rise to the setting of warning limits and rules concerning the number of consecutive observations that are increasing or decreasing in value. Page's CUSUM scheme uses a positive (negative) drift when defining the CUSUM statistics for detecting an increase (decrease) in mean. Crosier's CUSUM scheme shrinks the CUSUM statistic instead of adding a drift. Both of these CUSUM procedures make good choices for detecting small but persisting shifts in the process mean. They provide similar ARL performance according to our simulation study in Sect. 18.5 and in *Li* [18.6]. The EWMA scheme proposed by *Lucas* and *Saccucci* [18.4] is also an effective method for detecting small and persisting shifts. Unlike the other two univariate CUSUM schemes, the EWMA does not require the user to specify the shift in mean that should be detected. The EWMA scheme could have better ARL performance than these other CUSUM schemes when the shift is smaller than the specified value. However, the choice of the weight parameter r in (18.4) is somewhat crucial.

18.2 Multivariate Quality Monitoring Schemes

Often, more than one quality characteristics is measured on each unit. We then model the observations as a sequence of independent $p \times 1$ random vectors $\boldsymbol{X}_1$, $\cdots$, $\boldsymbol{X}_n$, $\cdots$ where each has the target mean value $\boldsymbol{a}$ and the same covariance matrix $\boldsymbol{\Sigma}$. If $\boldsymbol{\Sigma}$ is unknown, it must be estimated from a long sequence of observations taken when the process is stable. Any out-of-control observations, detected by a T^2 chart, are deleted. The sample covariance matrix is calculated from the reduced data set. Usually only one data cleaning stage is conducted. As in most of the literature, we will describe the multivariate monitoring schemes in terms of known covariance matrix $\boldsymbol{\Sigma}$. Most of the multivariate schemes discussed below involve a constant k and a positive control limit h.

The literature on multivariate process control includes [18.7–11]. Further discussion is given in [18.12], [18.13, Chapt. 10], and [18.14, Sects. 5.6 and 8.6].

To illustrate what might be the typical behavior of the various multivariate statistics, we generated a sequence of 100 bivariate normal observations with $\boldsymbol{\Sigma} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$. Our process mean shifted from (0,0) to (0,1) after the 40th observation. This is a moderate shift of one standard deviation in one of the two components.

In this section, we apply each monitoring scheme to that single bivariate sequence of observations. In Sect. 18.5, we look at various other choices and perform simulation studies to determine the ARL under a range of alternatives.

18.2.1 Multivariate T^2 Chart

The traditional T^2 chart reduces each multivariate observation to a scalar by defining

$$T_n^2 = (\boldsymbol{X}_n - \boldsymbol{a})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{X}_n - \boldsymbol{a}) \,. \tag{18.5}$$

The multivariate T^2 scheme signals a shift in mean when T_n^2 first exceeds a specified level h. That is, a change in mean is signaled when

$$T_n^2 \geq h \,.$$

The usual upper control limit is the upper χ^2 percentile $\chi_p^2(\alpha)$ with $\alpha = 0.01$. If the estimated covariance matrix $\boldsymbol{S}$ is based on a relatively small number of observations K, then the appropriate upper control limit is

$$\mathrm{UCL} = \frac{p(K-1)}{K-p} F_{p,K-p}(\alpha) \,.$$

As in the case with its univariate counterpart, the Shewhart X-bar chart, the multivariate T^2 procedure is not very sensitive to small or moderate shifts from the target $\boldsymbol{a}$ because it is only based on the most recent observation and has no *memory* of previous observations.

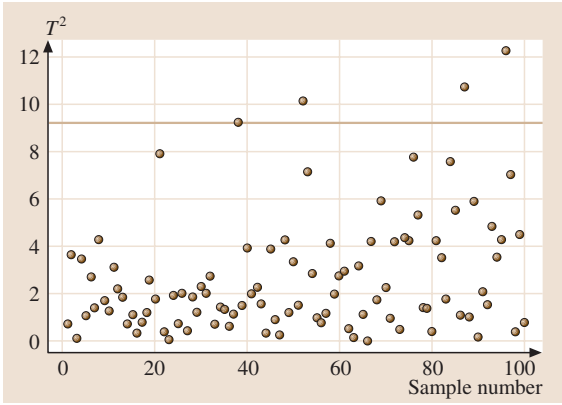

Fig. 18.6 T^2 chart with a 99% control limit using our generated bivariate data

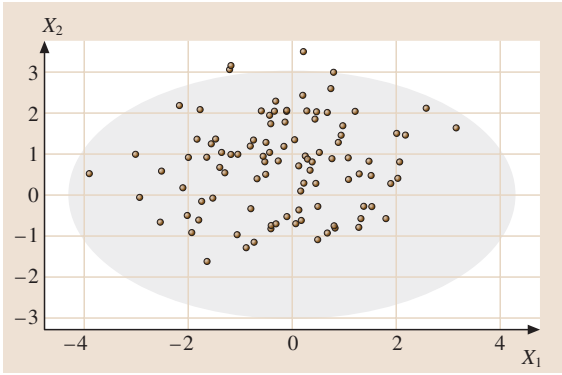

Fig. 18.7 The T^2 99% control ellipse using our generated bivariate data

Figure 18.6 illustrates the traditional multivariate T^2 scheme applied to our generated bivariate normal observations. As shown in the graph, there is a false alarm at $n = 38$ and the first correct detection of a shift is at $n = 52$. The shift from (0,0) to (1,0) at $n = 40$ is not detected particularly quickly and persistently by the T^2 chart with $\boldsymbol{a} = (0, 0)$.

When there are only two important variables, it is more informative to plot the individual 2×1 vectors $\boldsymbol{X}_n$ with the control ellipse in an ellipse format chart for T^2 statistics. The 99% control ellipse is

$$T_n^2 = (\boldsymbol{X}_n - \boldsymbol{a})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{X}_n - \boldsymbol{a}) \le \chi_2^2(0.01) \, .$$

The generated bivariate normal data and the 99% control ellipse are shown in Fig. 18.7. There is one false detection outside the control ellipse at $n = 38$ and a few correct detections of the shift in mean at $n = 52$, 87 and 96.

T^2 Control Charts Based on Means

Often, rather than using a single observation, a control chart is based on points that correspond to the mean of a small sample of size m. We still assume that the population is $p-$variate normal with mean vector $\boldsymbol{0}$ and covariance matrix $\boldsymbol{\Sigma}$. However, because of the central limit effect that sample means are more normal than individual observations, the normality of the population is not as crucial. It is now assumed that each random vector of observations from the process is independently distributed as $N_p(\boldsymbol{0}, \boldsymbol{\Sigma})$. We proceed differently, when the sampling procedure specifies that $m > 1$ units be selected, at the same time, from the process. From the first sample, we determine its sample mean $\overline{\boldsymbol{X}}_1$ and covariance matrix $\mathbf{S}_1$. When the population is normal, these two random quantities are independent.

Starting with K samples, where the the $j-$th sample has mean vector $\overline{\boldsymbol{X}}_j$ and covariance matrix $\mathbf{S}_j$, the estimator of the population mean vector $\boldsymbol{\mu}$ is the overall sample mean

$$\overline{\overline{\boldsymbol{X}}} = \frac{1}{K} \sum_{j=1}^{K} \overline{\boldsymbol{X}}_j \, .$$

The sample covariances from the n samples can be combined to give a single estimate (called $\boldsymbol{S}_{\text{pooled}}$) of $\boldsymbol{\Sigma}$ as

$$\mathbf{S} = \frac{1}{K} (\mathbf{S}_1 + \mathbf{S}_2 + \cdots + \mathbf{S}_K) \, ,$$

where $(mK - K)\mathbf{S}$ is independent of each $\overline{\boldsymbol{X}}_j$ and therefore of their mean $\overline{\overline{\boldsymbol{X}}}$. That is, we can now estimate $\boldsymbol{\Sigma}$ internally from the data collected in any given period. These estimators are combined into a single estimator with a large number of degrees of freedom.

T^2 Chart When Means Are Plotted. When the chart is based on the sample mean $\overline{\boldsymbol{X}}_n$ of m observations rather than a single observation, the values of

$$T_n^2 = m(\overline{\boldsymbol{X}}_n - \boldsymbol{a})' \boldsymbol{\Sigma}^{-1} (\overline{\boldsymbol{X}}_n - \boldsymbol{a})$$

are plotted for $n = 1, 2, \cdots$, where the

$$\text{UCL} = \chi_p^2(0.01)$$

or some other upper percentile of the chi-square distribution with p degrees of freedom.

Ellipse Format Chart. The control ellipse, expressed in terms of the sample mean $\overline{\boldsymbol{X}}_n$ of m observations, is

$$(\overline{\boldsymbol{x}} - \boldsymbol{a})' \boldsymbol{\Sigma}^{-1} (\overline{\boldsymbol{x}} - \boldsymbol{a}) \le \chi_p^2(0.01)/m \, ,$$

where $\chi_p^2(0.01)/m$, on the right-hand side, can be replaced by some other upper percentile. For $p = 2$, a graph of the ellipse is usually presented for visual inspection of the data.

18.2.2 CUSUM of T_n (COT) Scheme

The cumulative sum of T, COT or CUSUM of T_n scheme, is the most direct extension of the multivariate T^2 chart to a CUSUM procedure. It forms a CUSUM of the scalar statistics T_n, where T_n^2 is defined in (18.5).

Let $S_0 \geq 0$ and $k > 0$ be specified constants. Iteratively define

$$S_n = \max(0,\, S_{n-1} + T_n - k), \quad \text{for} \quad n = 1, 2, \cdots , \tag{18.6}$$

where k is a specified positive constant.

The COT scheme signals when it surpasses a specified level h. That is, the COT scheme signals a change in mean when

$$S_n \geq h \, .$$

Figure 18.8 illustrates the performance of the CUSUM of T. As we can see, the COT scheme does show an increasing trend which indicates a shift in the sequence mean, while the multivariate T^2 chart does not persistently indicate a change because the shift is not large enough.

18.2.3 Crosier's Multivariate CUSUM Scheme

Crosier [18.11] also generalized his univariate CUSUM scheme to the multivariate setting. Crosier's multivariate statistic starts at $\boldsymbol{S}_0 = \boldsymbol{0}$. With a specified constant k, iteratively define

$$\boldsymbol{S}_n = \begin{cases} \boldsymbol{0} & \text{if } C_n \leq k \\ (\boldsymbol{S}_{n-1} + \boldsymbol{X}_n - \boldsymbol{a})(1 - k/C_n) & \text{otherwise,} \end{cases} \tag{18.7}$$

where

$$C_n = [(\boldsymbol{S}_{n-1} + \boldsymbol{X}_n - \boldsymbol{a})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{S}_{n-1} + \boldsymbol{X}_n - \boldsymbol{a})]^{1/2} \, .$$

For a specified constant h, Crosier's multivariate scheme signals a shift in mean from $\boldsymbol{a}$ when

$$Y_n = (\boldsymbol{S}_n' \boldsymbol{\Sigma}^{-1} \boldsymbol{S}_n)^{1/2} \geq h \, .$$

Figure 18.9 illustrates the performance of Crosier's multivariate CUSUM scheme with the generated bivariate normal data. In the plot, a increasing trend was shown shortly after the shift in process occurs at the 41st observation. Therefore, the shift is detected faster than it is detected by the COT scheme.

18.2.4 Multivariate EWMA Scheme [MEWMA(r)]

In multivariate settings, *Lowry* and *Woodall* [18.15] proposed a natural extension of the univariate EWMA scheme. Starting with $\boldsymbol{Z}_0 = \boldsymbol{0}$, the multivariate EWMA statistic $\boldsymbol{Z}_n$ is defined, iteratively, by

$$\boldsymbol{Z}_n = \mathbf{R}(\boldsymbol{X}_n - \boldsymbol{a}) + (\mathbf{I} - \mathbf{R})\boldsymbol{Z}_{n-1} \quad \text{for} \quad n = 1, 2, \cdots , \tag{18.8}$$

where the weight matrix $\mathbf{R} = diag(r_1, \cdots , r_p), 0 \leq r_j \leq 1$, $j = 1, \cdots , p$. That is, Lowry and Woodall specialize

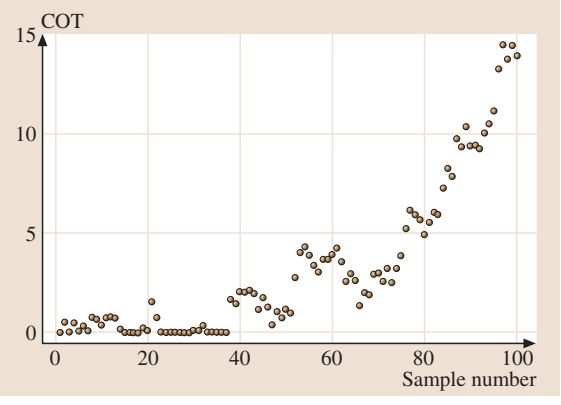

Fig. 18.8 CUSUM of T statistics using our generated bivariate data

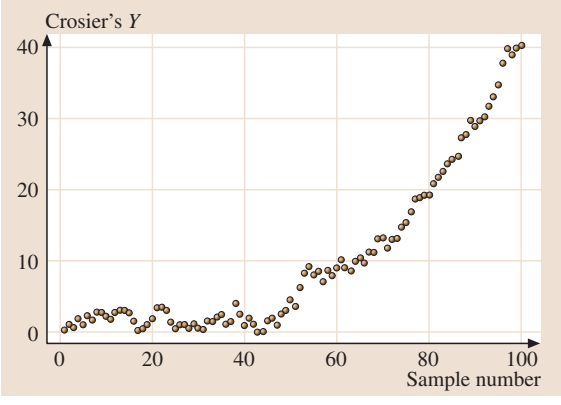

Fig. 18.9 Crosier's multivariate CUSUM statistics Y_n applied to our generated normal data

Index	x_1	x_2	x_3	x_4	x_5	x_6
1	−0.12	0.36	0.40	0.25	1.37	−0.13
2	−0.60	−0.35	0.04	−0.28	−0.25	−0.15
3	−0.13	0.05	0.84	0.61	1.45	0.25
4	−0.46	−0.37	0.30	0.00	−0.12	−0.25
5	−0.46	−0.24	0.37	0.13	0.78	−0.15
6	−0.46	−0.16	0.07	0.10	1.15	−0.18
7	−0.46	−0.24	0.13	0.02	0.26	−0.20
8	−0.13	0.05	−0.01	0.09	−0.15	−0.18
9	−0.31	−0.16	−0.20	0.23	0.65	0.15
10	−0.37	−0.24	0.37	0.21	1.15	0.05
11	−1.08	−0.83	−0.81	0.05	0.21	0.00
12	−0.42	−0.30	0.37	−0.58	0.00	−0.45
13	−0.31	0.10	−0.24	0.24	0.65	0.35
14	−0.14	0.06	0.18	−0.50	1.25	0.05
15	−0.61	−0.35	−0.24	0.75	0.15	−0.20
16	−0.61	−0.30	−0.20	−0.21	−0.50	−0.25
17	−0.84	−0.35	−0.14	−0.22	1.65	−0.05
18	−0.96	−0.85	0.19	−0.18	1.00	−0.08
19	−0.90	−0.34	−0.78	−0.15	0.25	0.25
20	−0.46	0.36	0.24	−0.58	0.15	0.25
21	−0.90	−0.59	0.13	0.13	0.60	−0.08
22	−0.61	−0.50	−0.34	−0.58	0.95	−0.08
23	−0.61	−0.20	−0.58	−0.20	1.10	0.00
24	−0.46	−0.30	−0.10	−0.10	0.75	−0.10
25	−0.60	−0.35	−0.45	0.37	1.18	−0.30
26	−0.60	−0.36	−0.34	−0.11	1.68	−0.32
27	−0.31	0.35	−0.45	−0.10	1.00	−0.25
28	−0.60	−0.25	−0.42	0.28	0.75	0.10
29	−0.31	0.25	−0.34	−0.24	0.65	0.10
30	−0.36	−0.16	0.15	−0.38	1.18	−0.10
31	−0.40	−0.12	−0.48	−0.34	0.30	−0.20
32	−0.60	−0.40	−0.20	0.32	0.50	0.10
33	−0.47	−0.16	−0.34	−0.31	0.85	0.60
34	−0.46	−0.18	0.16	0.01	0.60	0.35
35	−0.44	−0.12	−0.20	−0.48	1.40	0.10
36	−0.90	−0.40	0.75	−0.31	0.60	−0.10
37	−0.50	−0.35	0.84	−0.52	0.35	−0.75
38	−0.38	0.08	0.55	−0.15	0.80	−0.10
39	−0.60	−0.35	−0.35	−0.34	0.60	0.85
40	0.11	0.24	0.15	0.40	0.00	−0.10
41	0.05	0.12	0.85	0.55	1.65	−0.10
42	−0.85	−0.65	0.50	0.35	0.80	−0.21
43	−0.37	−0.10	−0.10	−0.58	1.85	−0.11
44	−0.11	0.24	0.75	−0.10	0.65	−0.10
45	−0.60	−0.24	0.13	0.84	0.85	0.15
46	−0.84	−0.59	0.05	0.61	1.00	0.20
47	−0.46	−0.16	0.37	−0.15	0.68	0.25
48	−0.56	−0.35	−0.10	0.75	0.45	0.20
49	−0.56	−0.16	0.37	−0.25	1.05	0.15
50	−0.25	−0.12	−0.05	−0.20	1.21	0.10

to cases where $\mathbf{R}$ is a diagonal matrix. This reduces to the situation where a univariate EWMA is applied to each individual component. When there is no a priori reason to weight the p quality characteristics differently, they further suggest the use of a common value $r_1 = \cdots = r_p = r$, where $0 \le r \le 1$ is a constant.

The MEWMA scheme signals if the scalar $\boldsymbol{Z}_n' \boldsymbol{\Sigma}^{-1} \boldsymbol{Z}_n$ is large. More particularly, for a specified constant h, the MEWMA scheme signals a shift in mean from $\boldsymbol{a}$ when

$$(\boldsymbol{Z}_n' \boldsymbol{\Sigma}^{-1} \boldsymbol{Z}_n)^{1/2} \ge h \,.$$

Analogous to the univariate case, the choice of weight matrix $\mathbf{R}$ has a considerable influence on the resulting ARL behavior. Usually a small value of r is selected to detect small shifts in each component of mean.

Figure 18.10 illustrates the performance of the MEWMA scheme evaluated for our generated bivariate normal data under two different choices of the constant r. Again, we see that the MEWMA scheme with a small value of r is more effective in detecting a small but persisting shift like the one in our generated data.

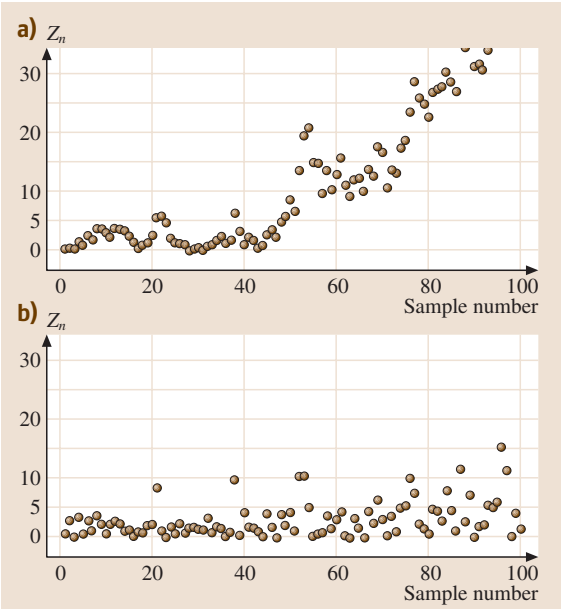

Fig. 18.10 MEWMA with $r = 0.1$ (a) and $r = 0.8$ (b), using our generated bivariate normal data

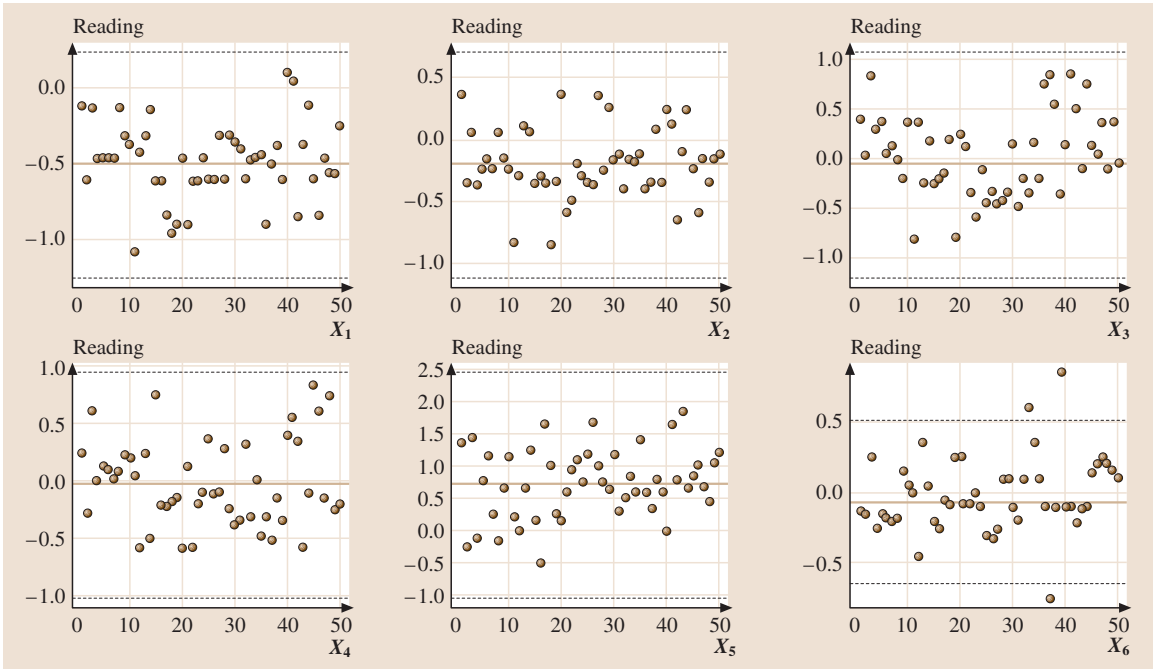

Fig. 18.11 Shewhart $X-$bar chart for each variable in automotive assembly data

18.3 An Application of the Multivariate Procedures

The data that we use to illustrate the various control charts is courtesy of Darek Ceglarek. He collected these measurements on the sheet metal assembly process as part of a study conducted with a major automobile manufacturer. *Ceglarek* and *Shi* [18.16] give more detail on the body assembly process. There are six variables of which four were measured when the car body was complete and two were measured on the underbody at an earlier stage of assembly.

All measurements were taken by sensors that recorded the deviation from the nominal value in millimeters:

x_1 = deviation at mid right-hand side, body complete,
x_2 = deviation at mid left-hand side, body complete,
x_3 = deviation at back right-hand side, body complete,
x_4 = deviation at back left-hand side, body complete,
x_5 = deviation at mid right-hand side of underbody,
x_6 = deviation at mid left-hand side of underbody.

The covariance matrix and the mean estimated from the first 30 observations are

$$\mathbf{S} = \begin{pmatrix} 0.0626 & 0.0616 & 0.0474 & 0.0083 & 0.0197 & 0.0031 \\ 0.0616 & 0.0924 & 0.0268 & -0.0008 & 0.0228 & 0.0155 \\ 0.0474 & 0.0268 & 0.1446 & 0.0078 & 0.0211 & -0.0049 \\ 0.0083 & -0.0008 & 0.0078 & 0.1086 & 0.0221 & 0.0066 \\ 0.0197 & 0.0228 & 0.0211 & 0.0221 & 0.3428 & 0.0146 \\ 0.0031 & 0.0155 & -0.0049 & 0.0066 & 0.0146 & 0.0366 \end{pmatrix}$$

$$\bar{\boldsymbol{x}} = (\,-0.5063 \quad -0.2070 \quad -0.0620 \quad -0.0317 \quad 0.6980 \quad -0.0650\,)'.$$

Figure 18.11 gives the Shewhart X-bar charts for each of the six variables. Except for two cases with the last variable measured at the left side of the underbody, all of the observations are within their 99% control limits.

Figures 18.12 and 18.13 show the multivariate T^2 chart and the CUSUM of T chart applied to the automotive assembly data. In the multivariate T^2 chart we can see that a few values between the 30th and 40th observations are out of the control limits. The CUSUM of T chart indicates a small and persistent shift at the end of the sequence, which is not detected by the multivariate T^2 chart.

The Crosier's CUSUM statistic illustrated in Fig. 18.14 also indicates a small and persistent shift, which is consistent with the CUSUM of T chart.

The multivariate EWMA schemes are illustrated in Fig. 18.15. There is a considerable difference in the appearance depending on the choice of the common weight r. In the plot of multivariate EWMA

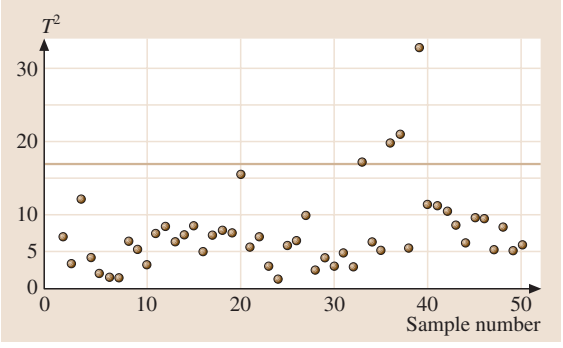

Fig. 18.12 Multivariate T^2 chart for automotive assembly data

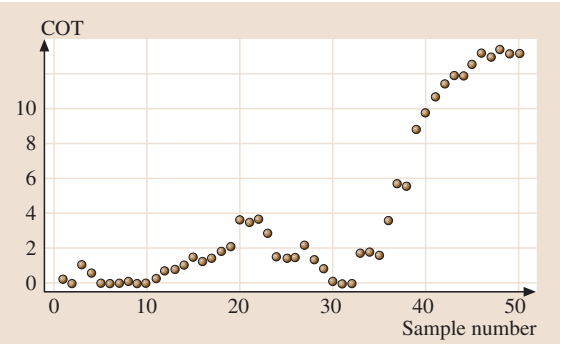

Fig. 18.13 CUSUM of T chart for automotive assembly data

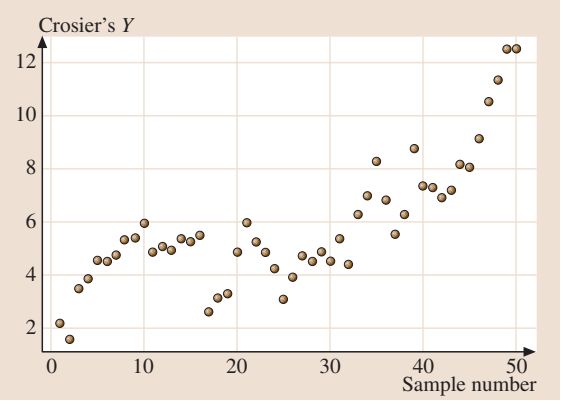

Fig. 18.14 Crosier's CUSUM statistic applied to automotive assembly data

statistics with $r = 0.1$, an increasing trend indicates an increase in the mean, which is consistent with the CUSUM of T chart and Crosier's CUSUM scheme. However, with $r = 0.8$, we can not see the increasing pattern in the plot for the multivariate EWMA statistics.

18.4 Comparison of Multivariate Quality Monitoring Methods

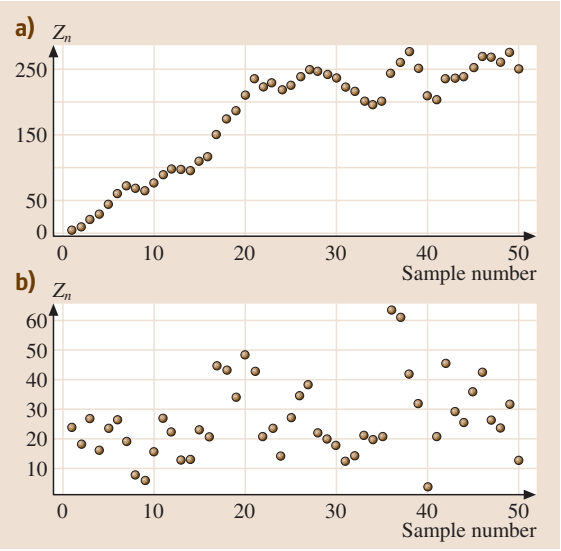

Fig. 18.15a,b Multivariate EWMA statistics with $r = 0.1$ (**a**) and $r = 0.8$ (**b**), using automotive assembly data

Several authors, including *Pignatiello* [18.17] and *Lowry* [18.18], have compared various multivariate monitoring procedures. To confirm their general conclusions, we use simulation to study the average run length (ARL) properties of the multivariate sequential schemes discussed above. Specifically, the ARL is defined as the average number of observations before the scheme gives a signal. If the shift did not occur at the beginning of the series of observations, it is common practice to use the steady-state ARL, which is the average additional run length after the shift occurs. An effective scheme should have a large ARL when the process is *in control* and a small ARL when the process is *out of control*. Usually, to compare different schemes, we can set the in-control ARLs to be nearly equal and compare the ARLs when there is a shift.

In our simulation, we generate bivariate normal observations with either $\boldsymbol{\Sigma} = \boldsymbol{I}$ (uncorrelated) or $\boldsymbol{\Sigma} = \begin{pmatrix} 1.0 & -0.6 \\ -0.6 & 1.0 \end{pmatrix}$. For each choice of $\boldsymbol{\Sigma}$ and

Table 18.1 ARL comparison with bivariate normal data (uncorrelated)

Shift	T^2	COT	Crosier	EW(0.1)	EW(0.4)	EW(0.8)
(0,0)	198.3	200.1	200.8	200.4	199.4	198.1
	2.8	2.7	2.9	2.8	2.8	2.8
(1,0)	44.3	20.5	9.5	9.8	13.3	29.6
	0.6	0.2	0.1	0.1	0.1	0.4
(0,-1)	42.9	20.0	9.5	9.8	13.2	28.7
	0.6	0.2	0.1	0.1	0.2	0.4
(0.7,0.7)	42.4	20.3	9.5	9.7	12.7	29.0
	0.6	0.2	0.1	0.1	0.2	0.4
(0.7,-0.7)	42.3	20.3	9.3	9.7	13.4	29.1
	0.6	0.2	0.1	0.1	0.2	0.4
(0.5,0)	119.7	82.4	29.0	27.3	54.2	96.4
	1.6	1.1	0.3	0.3	0.7	1.3
(0.4,-0.4)	113.9	82.1	29.0	27.5	52.8	90.7
	1.6	1.1	0.3	0.3	0.7	1.3
(2,0)	6.9	4.7	4.0	4.3	3.5	4.8
	0.1	0.0	0.0	0.0	0.0	0.1
(1.4,-1.4)	6.7	4.7	4.0	4.3	3.5	4.7
	0.1	0.0	0.0	0.0	0.0	0.1

Table 18.2 ARL comparison with bivariate normal data (correlated)

Shift	T^2	COT	Crosier	EW(0.1)	EW(0.4)	EW(0.8)
(0,0)	201.1	206.5	202.3	203.0	202.7	202.6
	2.8	2.7	3.0	2.8	2.8	2.8
(0.8,0)	42.7	20.0	9.4	9.8	13.1	28.3
	0.6	0.2	0.1	0.1	0.2	0.4
(0,−0.8)	44.6	20.6	9.7	9.9	13.4	29.8
	0.6	0.2	0.1	0.1	0.2	0.4
(0.4,0.4)	43.3	20.0	9.5	9.8	13.1	28.8
	0.6	0.2	0.1	0.1	0.2	0.4
(0.4,−0.4)	115.3	81.8	28.5	26.9	54.2	97.2
	1.6	1.1	0.3	0.3	0.7	1.3
(0.4,0)	114.3	84.1	28.8	27.0	52.8	95.0
	1.6	1.1	0.3	0.3	0.7	1.4
(0.2,−0.2)	172.3	158.2	84.2	71.7	115.5	158.7
	2.4	2.1	1.1	1.0	1.7	2.2
(1.6,0)	7.0	4.7	4.0	4.3	3.5	4.7
	0.1	0.0	0.0	0.0	0.0	0.1
(0.9,−0.9)	41.8	20.4	9.5	10.1	14.4	28.5
	0.6	0.2	0.1	0.1	0.2	0.4

shift in mean, series of observations were generated and the multivariate statistics were calculated until a shift is signaled. This procedure was repeated 5000 times and we calculate the ARL and the estimated standard error of ARL for each scheme. Table 18.1 and 18.2 show the results of our simulation, where the estimated standard errors are given in smaller type.

From our simulation and existing literature, we conclude that, due to the fact that the value of the T^2 statistic only depends on the most current observation, it is not sensitive to small and moderate shifts in the mean of a process even if the shift is persistent. By taking the CUSUM of T, The COT procedure has ARL performance that is significantly improved over that of the T^2 chart. The ARL of Crosier's scheme is considerably better than that of the COT scheme. The performance of multivariate EWMA schemes depend heavily on the value of the weight parameters. If the weight is appropriately selected, the multivariate EWMA will have very good ARL performance which is comparable with Crosier's scheme.

18.5 Control Charts Based on Principal Components

The first two sample principal components concentrate the sample variability. Starting with a sample $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_K$, of size K, collected when the process is stable, the first sample principal component is the linear combination with values, $y_j = \boldsymbol{a}'(\boldsymbol{x}_j - \overline{\boldsymbol{x}})$ that has maximum sample variance among all choices with $\boldsymbol{a}'\boldsymbol{a} = 1$. The second sample principal component is the linear combination, having values $\boldsymbol{b}'(\boldsymbol{x}_j - \overline{\boldsymbol{x}})$, that has maximum variance among all linear combinations with $\boldsymbol{b}'\boldsymbol{b} = 1$ and that have zero correlation with the first principal component. The third sample principal component is the linear combination with maximum sample variance, subject to being uncorrelated with each of the first two principal components and having coefficient vector of length one. See *Johnson* and *Wichern* [18.14] for a thorough description of principal components.

The coefficients that provide the maximum variance are the eigenvectors $\boldsymbol{e}$ of the sample covariance matrix $\mathbf{S}$. That is, they are the solutions of

$$\mathbf{S}\,\boldsymbol{e} = \lambda\,\boldsymbol{e}$$

with the $\boldsymbol{e}$ normalized so that $\boldsymbol{e}'\boldsymbol{e} = 1$. There are p solutions $(\lambda_i, \boldsymbol{e}_i)$ where $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_p$. The first sample principal component

$$\hat{y}_{1j} = \boldsymbol{e}_1'(\boldsymbol{x}_j - \overline{\boldsymbol{x}})$$

has the maximum possible variance λ_1. More generally, the k-th principal component

$$\hat{y}_{kj} = \boldsymbol{e}_k'(\boldsymbol{x}_j - \overline{\boldsymbol{x}}), \quad k = 1, 2, \cdots, p$$

has sample variance λ_k.

If the process is stable over time, then the values of the first two principal components should be stable. Conversely, if the principal components remain stable over time, the common effects which influence the process are likely to remain constant.

Through an example, we introduce a two-part monitoring procedure based on principal components.

18.5.1 An Application Using Principal Components

Today, with electronic and other automated methods of data collection, major chemical and drug companies regularly measure over 100 different process variables such as temperature, pressure, concentrations and weights at various positions along the production process. Even with 11 variables to monitor, there are 55 possible pairs of variables for which ellipse format charts could be created. Consequently, we need to consider an alternative approach that both produces visual displays of important quantities and still has the sensitivity to detect special causes. Here we introduce a two-part multivariate quality control procedure that is widely applied when a large number of variables are being monitored [18.19].

The first part is an ellipse format chart to monitor the first two principal components. The second part is a T^2 chart based on the remaining principal components.

If the p variables have quite different variances, it is usual to standardize the variables before finding the principal components. This is equivalent to extracting the eigenvectors from the correlation matrix $\mathbf{R}$. Here we skip that step because the variables are comparable.

The values of the first and second sample principal components for the $n-$th observation are

$$\hat{y}_{1n} = \boldsymbol{e}_1'(\boldsymbol{x}_n - \overline{\boldsymbol{x}}) ,$$
$$\hat{y}_{2n} = \boldsymbol{e}_2'(\boldsymbol{x}_n - \overline{\boldsymbol{x}}) .$$

The first part of the multivariate quality control procedure is to construct a ellipse format chart for the pairs of values $(\hat{y}_{1n}, \hat{y}_{2n})$, $n = 1, 2, \ldots$.

Recall that the variance of the first principal component $\hat{y}_1$ is λ_1, that of the second principal component $\hat{y}_2$ is λ_2, and the two are uncorrelated. Consequently, the control format ellipse reduces to the collection of possible values $(\hat{y}_1, \hat{y}_2)$ such that

$$\frac{\hat{y}_1^2}{\lambda_1} + \frac{\hat{y}_2^2}{\lambda_2} \le \chi_2^2(\alpha) .$$

Chart 1: The elliptical control region for the first two principal components

Refer to the automotive assembly data on $p = 6$ variables. We base our estimate $\mathbf{S}$ of $\boldsymbol{\Sigma}$ on the first 30 stable observations. A computer calculation gives the eigenvalues and eigenvectors in Table 18.3.

The 99% ellipse format chart for the first two principal components

$$\frac{\hat{y}_1^2}{\lambda_1} + \frac{\hat{y}_2^2}{\lambda_2} \le \chi_2^2(0.01)$$

is shown in Fig. 18.16 along with the pairs of values of the principal components for the first 30 observations as well as the additional cases. There are no points out of control.

Special causes may still produce shocks to the system not apparent in the values of the first two principal components and a second chart is required.

Chart 2: A T^2 chart for the remaining principal components

For the 30 stable observations, the approximation to $\boldsymbol{x}_j - \overline{\boldsymbol{x}}$ by the first two principal components has the form $\hat{y}_{1j}\boldsymbol{e}_1 + \hat{y}_{1j}\boldsymbol{e}_2$ [18.14]. This leaves an unexplained component of the j-th deviation $\boldsymbol{x}_j - \overline{\boldsymbol{x}}$. Namely,

$$\boldsymbol{x}_j - \overline{\boldsymbol{x}} - \hat{y}_{1j}\boldsymbol{e}_1 - \hat{y}_{2j}\boldsymbol{e}_2 .$$

For each j, this unexplained component is perpendicular to both of the eigenvectors $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$. Consequently,

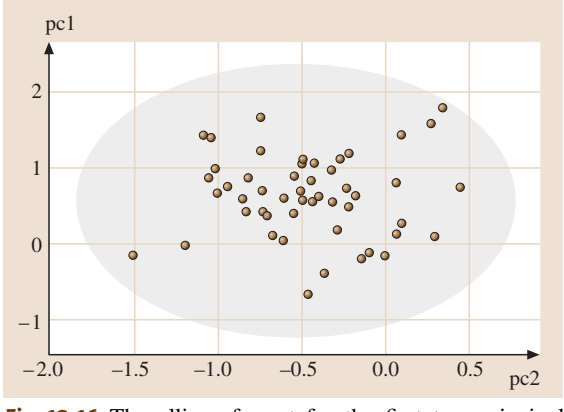

Fig. 18.16 The ellipse format for the first two principal components – automotive data

Table 18.3 Eigenvectors and eigenvalues from the 30 stable observations

	e_1	e_2	e_3	e_4	e_5	e_6
x_1	0.1193	0.4691	0.0752	0.2906	0.2672	0.7773
x_2	0.1295	0.4576	0.2508	0.6237	0.0366	−0.5661
x_3	0.1432	0.7170	−0.1161	−0.6323	−0.1746	−0.1471
x_4	0.0964	0.0529	−0.9555	0.2568	0.0619	−0.0715
x_5	0.9677	−0.2312	0.0700	−0.0641	0.0322	−0.0030
x_6	0.0517	0.0135	−0.0078	0.2380	−0.9444	0.2204
λ_i	0.3544	0.1864	0.1076	0.0972	0.0333	0.0088

it is perpendicular to its approximation $\hat{y}_{1j}\boldsymbol{e}_1 + \hat{y}_{1j}\boldsymbol{e}_2$, which implies that the approximation and unexplained component have zero sample covariance.

Seen another way, let $\mathbf{E} = (\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_p)$ be the orthogonal matrix whose columns are the eigenvectors of $\mathbf{S}$. The orthogonal transformation of the unexplained part is

$$\mathbf{E}'(\boldsymbol{x}_j - \overline{\boldsymbol{x}} - \hat{y}_{1j}\boldsymbol{e}_1 - \hat{y}_{2j}\boldsymbol{e}_2) = \begin{pmatrix} 0 \\ 0 \\ \hat{y}_{3j} \\ \cdot \\ \cdot \\ \cdot \\ \hat{y}_{pj} \end{pmatrix} .$$

The first two components are always zero so we base the T^2 chart on the values of the last $p-2$ principal components. Because the

$$\text{sample variance } (\hat{y}_{ij}) = \lambda_i \text{ for } i = 1, 2, 3, \cdots, p$$

and the principal components have zero sample covariances, the T^2 based on the original quantities $\boldsymbol{x}_j - \overline{\boldsymbol{x}} - \hat{y}_{1j}\boldsymbol{e}_1 - \hat{y}_{2j}\boldsymbol{e}_2$ is equivalent to the one based on the values

$$T_j^2 = \frac{\hat{y}_{3j}^2}{\lambda_3} + \frac{\hat{y}_{4j}^2}{\lambda_4} + \cdots + \frac{\hat{y}_{pj}^2}{\lambda_p} .$$

Because the coefficients of the linear combinations, $\boldsymbol{e}_i$, are also estimates, the principal components do not have a normal distribution even when the underlying population is normal. Consequently, it is customary to use the large-sample approximation to the upper control limit, UCL = $\chi^2_{p-2}(\alpha)$.

This T^2 statistic can be based on high-dimensional data. When $p = 20$ variables are measured, it concerns the 18-dimensional space perpendicular to the first two eigenvectors. Still, it is reported as highly effective in picking up special causes.

Refer the automobile assembly data. The quality ellipse for the first two principal components was shown in Fig. 18.16. To illustrate the second step of the two step monitoring procedure, we create the chart for the other variables. Since $p = 6$, this 99% chart is based on $6-2 = 4$ dimensions and the upper control limit is $\chi^2_4(0.01) = 13.28$. We plot the time sequence of values

$$T_n^2 = \frac{\hat{y}_{3n}^2}{\lambda_3} + \frac{\hat{y}_{4n}^2}{\lambda_4} + \frac{\hat{y}_{5n}^2}{\lambda_5} + \frac{\hat{y}_{6n}^2}{\lambda_6} .$$

The T^2 chart is shown as Fig. 18.17.

Something has likely happened at the 33rd, 36th, 37th and the 39th observations. For the 39th observation, the values of the last principal components are -0.2712, -0.2105, 0.8663, -0.2745, respectively. The value of $\hat{y}_{5,39} = 0.8663$ is particularly large, with reference to the coefficient vector $\boldsymbol{e}_5$ in Table 18.3, we see that the fifth principal component is essentially X_6. From the data and the univariate Shewhart X-bar charts in Fig. 18.11, we see that the mean of the last variable has increased.

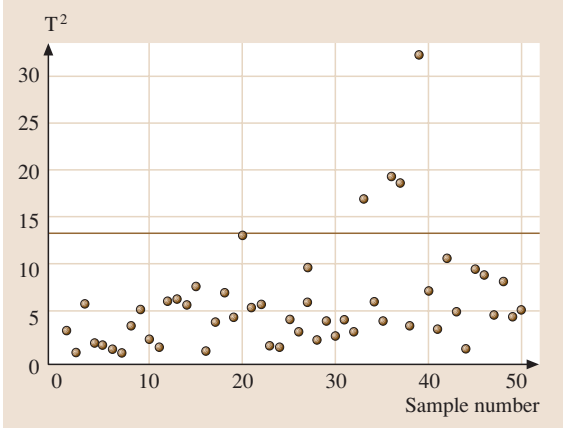

Fig. 18.17 T^2 chart based on the last four principal components – automotive data

In some applications in the pharmaceutical industry hundreds of variables are monitored. Then, the space orthogonal to the first few principal components has dimension greater than 100 and some the eigenvalues are very small. An alternative approach [18.19], which avoids the difficulty of dividing by very small numbers, has been successful applied. For each of the K stable observations, take the sum of squares of its unexplained component

$$d^2_{\perp j} = (\boldsymbol{x}_j - \overline{\boldsymbol{x}} - \hat{y}_{1j}\boldsymbol{e}_1 - \hat{y}_{2j}\boldsymbol{e}_2)' \times (\boldsymbol{x}_j - \overline{\boldsymbol{x}} - \hat{y}_{1j}\boldsymbol{e}_1 - \hat{y}_{2j}\boldsymbol{e}_2) .$$

Note that, by inserting $\mathbf{E}\mathbf{E}' = \mathbf{I}$, we also have

$$d^2_{\perp j} = \left(\boldsymbol{x}_j - \overline{\boldsymbol{x}} - \hat{y}_{1j}\boldsymbol{e}_1 - \hat{y}_{2j}\boldsymbol{e}_2\right)' \mathbf{E}\mathbf{E}' \times \left(\boldsymbol{x}_j - \overline{\boldsymbol{x}} - \hat{y}_{1j}\boldsymbol{e}_1 - \hat{y}_{2j}\boldsymbol{e}_2\right) = \sum_{k=3}^{p} \hat{y}^2_{kj}$$

which is just the sum of squares of the neglected principal components.

Using either form, the $d^2_{\perp n}$ are plotted versus n to create a control chart. The lower control limit is zero and the upper limit is set by approximating the distribution of $d^2_{\perp n}$ as a constant c times a chi-square random variable with ν degrees of freedom. The constant c and degrees of freedom ν are chosen to match the sample mean and variance of the $d^2_{\perp j}$, $j = 1, \cdots, K$. In particular, setting

$$\overline{d^2_{\perp}} = \frac{1}{K}\sum_{j=1}^{K} d^2_{\perp j} = c\,\nu ,$$

$$s_{dd} = \frac{1}{K}\sum_{j=1}^{K}\left(d^2_{\perp j} - \overline{d^2_{\perp}}\right)^2 = 2c^2\nu ,$$

we determine that

$$c = \frac{s_{dd}}{2\overline{d^2_{\perp}}} \quad \text{and} \quad \nu = 2\frac{(\overline{d^2_{\perp}})^2}{s_{dd}} .$$

The upper control limit is then $c\chi^2_\nu(\alpha)$, where $\alpha = 0.05$ or 0.01.

We remark that this two-step procedure can be made more sensitive by using, for instance, Crosier's scheme.

18.6 Difficulties of Time Dependence in the Sequence of Observations

We must include a warning that pertains to the application of any of the quality monitoring procedures we have discussed. They are all based on the assumption that the observations are independent. Most often, especially with automated collection procedures, observations may be taken close together in time or space. This can produce a series of observations that are not independent but correlated in time.

We emphasize that the methods described in this chapter are based on the assumption that the multivariate observations $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n$ are independent of one another. The presence of even a moderate amount of time dependence among the observations can cause serious difficulties for monitoring procedures.

One common and simple univariate model that usually captures much of the time dependence is a first-order autoregressive process (AR(1))

$$X_n - \mu = \phi\,(X_{n-1} - \mu) + \varepsilon_n ,$$

where $-1 < \phi < 1$. The ε_n are independent, mean-zero, shocks all having the same variance σ^2_ε. The AR(1) model relates the observation at time n, to the observation at time $n-1$, through the coefficient ϕ. The name autoregressive model comes from the fact that the model looks like a regression model with X_n as the dependent variable and the previous value X_{n-1} as the independent variable.

Under normality of errors, if $\phi = 0$, the autoregressive model implies that the observations are independent. It is not at all unusual in practice to find values of ϕ as high as 0.3 or 0.4.

The AR(1) model, implies that all of the X_n have the same variance $\sigma^2_X = \sigma^2_\varepsilon / (1-\phi^2)$. As shown in *Johnson* and *Langeland* [18.20], if the sample variance s^2 is calculated from a long series of adjacent observations, s^2 will closely approximate σ^2_X. That is, the correct variance is being estimated.

We first consider the effects of dependence in the context of the X-bar chart. When the individual observations X_n are plotted, with limits set at $\pm 3\sigma_X$, we still have

$$P(\,|\,X_n - \mu\,|\,> 3\sigma_X\,) = 0.0027 .$$

The observations are correlated but we expect the same number of false alarms.

In this sense, the X-bar chart is robust with respect to dependence. If the points being plotted, $\overline{X}_n$, correspond

Table 18.4 Probability of false alarms when the process is in control. Normal populations and X-bar chart

	$P\left(\lvert\overline{X}_n - \mu\rvert > 3\sigma_X/\sqrt{m}\right)$				
	$\phi = -0.6$	$\phi = -0.3$	$\phi = 0$	$\phi = 0.3$	$\phi = 0.6$
$m = 1$	0.0027	0.0027	0.0027	0.0027	0.0027
$m = 3$	0.0000	0.0003	0.0027	0.0147	0.0484
$m = 5$	0.0000	0.0002	0.0027	0.0217	0.0971

Table 18.5 The estimated ARL for Page's CUSUM when the process is in control. Normal populations

	$\phi = -0.4$	$\phi = -0.2$	$\phi = -0.1$	$\phi = 0$	$\phi = 0.1$	$\phi = 0.2$	$\phi = 0.4$
$k = 0.5,\ \ h = 4.24$	9549.0	884.0	390.5	206.1	121.8	82.6	43.9
$k = 0.75,\ \ h = 2.96$	2901.2	706.3	356.2	204.1	128.5	88.5	48.1

Table 18.6 The h value to get in-control ARL ≈ 200, $k = 0.5$. Page's CUSUM

	$\phi = -0.6$	$\phi = -0.3$	$\phi = -0.1$	$\phi = 0$	$\phi = 0.1$	$\phi = 0.3$	$\phi = 0.6$
h	2.34	2.94	3.77	4.24	4.76	6.38	10.6
in-control ARL	214.6	205.5	213.2	206.1	202.7	205.8	210.3

to the sample mean of m adjacent observations, then

$$\sigma^2_{\overline{X}} = \mathrm{Var}(\overline{X}_n) = \frac{\sigma^2_X}{m}\left(1 + \sum_{j=1}^{m-1}\frac{m-j}{m}\phi^j\right).$$

In this case, the probability of a false alarm when the limits are set at $3\sigma_X/\sqrt{m}$, is

$$P\left(\lvert\overline{X}_n - \mu\rvert > 3\frac{\sigma_X}{\sqrt{m}}\right) = P\left(\left|\frac{\overline{X}_n - \mu}{\sigma_{\overline{X}}}\right| > 3\frac{\sigma_X/\sqrt{m}}{\sigma_{\overline{X}}}\right).$$

There are some dramatic changes from the nominal value of the probability of a false alarm when the process is in control. Some values are given in Table 18.4.

The consequences of dependence are much more severe on the CUSUM statistic. *Johnson* and *Bagshaw* [18.21] present some limiting results for the distribution of ARL when the centering values are selected so the contribution of the n-th observation has mean zero for every n. The presence of dependence greatly alters the ARL. Essentially, this occurs because the CUSUM is a smooth function of the stochastic process defined at $n = 0, 1, \cdots$ by

$$n^{-1/2}S_n = n^{-1/2}\sum_{i=1}^{n}(X_i - \mu).$$

For large fixed n, $n^{-1/2}S_n$ is normal with variance approaching $\sigma^2_X(1+\phi)/(1-\phi)$. This is quite different from σ^2_X, which is the value if we ignore the dependence in the model.

Table 18.5 gives the estimated ARL using Page's CUSUM procedure, for in-control normal populations, for a few values of ϕ. The h values in the CUSUM scheme are chosen carefully so that the in-control ARL at $\phi = 0$ is around 200. The ARL values are based on 5000 trials and apply to the situation where the dependence is not noticed when $\phi \neq 0$, but that variance is estimated by the usual formula using a long series of consecutive observations.

We see dramatic reductions in ARL even for small positive values of ϕ. To be able to get the desired in-control ARL when there is time dependence in the sequence of observations, the h values in the control schemes have to be changed. Table 18.6 presents the h values obtained from simulation to get in-control ARL near 200, for different values of ϕ, using Page's CUSUM scheme when the dependence is not noticed.

In the context of multivariate procedures, dependence can often be represented as a multivariate first-order autoregressive model. Let the $p \times 1$ random vector $\boldsymbol{X}_t$ follow the multivariate AR(1) model

$$\boldsymbol{X}_n - \boldsymbol{\mu} = \boldsymbol{\Phi}(\boldsymbol{X}_{n-1} - \boldsymbol{\mu}) + \varepsilon_n \qquad (18.9)$$

where the ε_n are independent and identically distributed with $E(\varepsilon_n) = \boldsymbol{0}$ and $\mathrm{Cov}(\varepsilon_n) = \boldsymbol{\Sigma}_\varepsilon$ and all of the eigenvalues of $\boldsymbol{\Phi}$ are between -1 and 1. Under this model

$$\mathrm{Cov}(\boldsymbol{X}_n, \boldsymbol{X}_{n-j}) = \boldsymbol{\Phi}^j\boldsymbol{\Sigma}_X,$$

$$\text{where } \boldsymbol{\Sigma}_X = \sum_{j=0}^{\infty}\boldsymbol{\Phi}^j\boldsymbol{\Sigma}_\varepsilon\boldsymbol{\Phi}'^j.$$

Table 18.7 The estimate in-control ARL using Crosier's multivariate scheme

	$\phi = -0.3$	$\phi = -0.2$	$\phi = -0.1$	$\phi = 0$	$\phi = 0.1$	$\phi = 0.2$	$\phi = 0.3$
$k = 0.5,\ \ h = 5.55$	4881.3	1295.2	454.8	203.6	105.5	65.6	43.2

The multivariate AR(1) model relates the observation at time n to the observation at time $n-1$ through the coefficient matrix Φ. Further, the autoregressive model says the observations are independent, under multivariate normality, if all the entries in the coefficient matrix Φ are 0.

As shown in *Johnson* and *Langeland* [18.20],

$$\overline{\boldsymbol{X}} \to^{P} \boldsymbol{\mu}\ ,$$

$$\boldsymbol{S} = \frac{1}{n-1}\sum_{t=1}^{n}(\boldsymbol{X}_t - \overline{\boldsymbol{X}})(\boldsymbol{X}_t - \overline{\boldsymbol{X}})' \to^{P} \boldsymbol{\Sigma}_X\ ,$$

where the arrow indicates convergence in probability.

Similar to the X-bar chart, the T^2 chart is robust with respect to dependence when individual observations are plotted. The probability of a false alarm at any time n is still

$$P\left[(\boldsymbol{X}_n - \boldsymbol{\mu})'\boldsymbol{\Sigma}_X^{-1}(\boldsymbol{X}_n - \boldsymbol{\mu}) \geq \chi_p^2(0.01)\right] = 0.01$$

since, under normality, $\boldsymbol{X}_n - \boldsymbol{\mu}$ has a $p-$variate normal distribution with mean $\boldsymbol{0}$ and covariance matrix $\boldsymbol{\Sigma}_X$. If the observations are independent, and the process is in control, the probability is 0.01 that a single observation will falsely signal that a change has occurred. The ARL is then $1/0.01 = 100$.

When the plotted points correspond to the sample mean of m adjacent observations, the situation is more complicated. To simplify our calculations, we consider the case where has $\boldsymbol{\Phi} = \phi\mathbf{I}$, where $|\phi| < 1$. Consider the multivariate T^2 chart where the process is considered to be in control if

$$m(\overline{\boldsymbol{X}}_n - \boldsymbol{\mu})'\Sigma_X^{-1}(\overline{\boldsymbol{X}}_n - \boldsymbol{\mu}) \leq \chi_p^2(0.01)\ .$$

If the observations are independent, and the process is in control, the probability is 0.01 that a single observation will falsely signal that a change has occurred. The ARL is $1/0.01 = 100$. Using $\chi_p^2(0.005)$ as limit gives an ARL of $1/0.005 = 200$ when the process is in control. However, if the observations are related by our simplified multivariate AR(1) model, the average of m adjacent observations has covariance matrix

$$\mathrm{Cov}(\overline{\boldsymbol{X}}_n) = \frac{1}{m}\boldsymbol{\Sigma}_X\left(1 + \sum_{j=1}^{m-1}\frac{m-j}{m}\phi^j\right)\ .$$

not $\boldsymbol{\Sigma}_X/m$, and this will cause some change from the nominal probability.

The multivariate CUSUM statistics are all based on the stochastic process, defined at $n = 1, 2, \cdots$ by

$$n^{-1/2}\boldsymbol{S}_n = n^{-1/2}\sum_{j=1}^{n}(\boldsymbol{X}_j - \boldsymbol{\mu})$$

whose covariance at time n

$$\mathrm{Cov}\left(n^{-1/2}\sum_{t=1}^{n}\boldsymbol{X}_t\right) \to (\boldsymbol{I} - \Phi)^{-1}\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_X(\boldsymbol{I} - \Phi')^{-1} - \boldsymbol{\Sigma}_X\ .$$

This can be considerably different from the covariance $\boldsymbol{\Sigma}_X$ used when dependence is ignored.

Table 18.7 provides the estimated ARL for in-control normal populations with covariance matrix $\boldsymbol{\Sigma} = \begin{pmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{pmatrix}$, using Crosier's CUSUM scheme. The ARL values are based on 5000 trials where the covariance matrix is estimated by $\boldsymbol{S}$ using a long series of consecutive observations.

For more details on large-sample approximations, including a limit for Crosiers statistic, see *Li* [18.6].

Based on the calculations above and consideration of other cases, we must emphasize that the independence assumption is crucial to all of the procedures we discussed that are based on cumulative sums. The results based on this assumption can be seriously misleading if the observations are, in fact, even moderately dependent.

The best approach, when dependence is identified as being present, is to fit a time-series model. Then, as suggested in *Bagshaw* and *Johnson* [18.22], the residuals can be monitored for a shift. In the univariate case a CUSUM statistic can be applied. See also *Hawkins* and *Olwell* [18.23], Section 9.3.

References

18.1 W. A. Shewhart: *Economic Control of Quality of Manufactured Product* (Van Nostrand, New York 1931)

18.2 E. S. Page: Continuous inspection schemes, Biometrika **41**, 100–115 (1954)

18.3 R. B. Crosier: A new two-sided cumulative sum quality control scheme, Technometrics **28**, 187–194 (1986)

18.4 J. M. Lucas, M. S. Saccucci: Exponentially weighted moving average control schemes: properties and enhancements, Technometrics **32**, 1–12 (1990)

18.5 D. C. Montgomery: *Introduction to Statistical Quality Control*, 4th edn. (Wiley, New York 2000)

18.6 Li, R. New Multivariate Schemes for Statistical Process Control, Dissertation, Department of Statistics, Univ. of Wisconsin (2004)

18.7 J. E. Jackson: Quality control methods for several related variables, Technometrics **1**, 359–377 (1959)

18.8 J. E. Jackson: Multivariate quality control, Commun. Stat. A **14**, 2657–2688 (1985)

18.9 N. Doganaksoy, J. Fulton, W. T. Tucker: Identification of out of control quality characteristics in multivariate manufacturing environment, Commun. Stat. A **20**, 2775–2790 (1991)

18.10 N. D. Tracy, J. C. Young, R. L. Mason: Multivariate quality control charts for individual observations, J. Qual. Technol. **24**, 88–95 (1992)

18.11 R. B. Crosier: Multivariate generalizations of cumulative sum quality-control schemes, Technometrics **30**, 291–303 (1988)

18.12 C. Fuchs, R. S. Kenett: *Multivariate Quality Control: Theory and Applications* (Marcel Dekker, New York 1998)

18.13 K. Yang, J. Trewn: *Multivariate Statistical Methods in Quality Management* (McGraw-Hill, New York 2004)

18.14 R. A. Johnson, D. W. Wichern: *Applied Multivariate Statistical Analysis* (Prentice Hall, Piscataway 2002)

18.15 C. A. Lowry, W. H. Woodall, C. W. Champ, S. E. Rigdon: A multivariate exponentially weighted moving average control chart, Technometrics **34**, 46–53 (1992)

18.16 D. Ceglarek, J. Shi: Dimensional variation reduction for automotive body assembly, Manuf. Rev. **8**, 139–154 (1995)

18.17 J. J. Pignatiello, G. C. Runger: Comparisons of multivariate CUSUM charts, J. Qual. Technol. **22**, 173–186 (1990)

18.18 C. A. Lowry, D. C. Montgomery: A review of multivariate control charts, IIE Trans. **27**, 800–810 (1995)

18.19 T. Kourti, J. F. MacGregor: Multivariate SPC methods for process and product monitoring, J. Qual. Technol. **28**, 409–428 (1996)

18.20 R. A. Johnson, T. Langeland: A linear combinations test for detecting serial correlation in multivariate samples. In: *Statistical Dependence*, Topics, ed. by H. Block et al. (Inst. Math. Stat. Mon. 1991) pp. 299–313

18.21 R. A. Johnson, M. Bagshaw: The effect of serial correlation on the performance of CUSUM tests, Technometrics **16**, 103–112 (1974)

18.22 M. Bagshaw, R. A. Johnson: Sequential procedures for detecting parameter changes in a time-series model, J. Am. Stat. Assoc. **72**, 593–597 (1977)

18.23 D. M. Hawkins, D. Olwell: *Cumulative Sum Charts and Charting for Quality Improvement* (Springer, New York 1998)