

Process Monitoring

In industrial plants, large numbers of process variables must be maintained within specified limits in order for the plant to operate properly. Excursions of key variables beyond these limits can have significant consequences for plant safety, the environment, product quality, and plant profitability. Earlier chapters have indicated that industrial plants rely on feedback and feedforward control to keep process variables at or near their set points. A related activity, *process monitoring*, also plays a key role in ensuring that the plant performance satisfies the operating objectives. In this chapter, we introduce standard monitoring techniques as well as newer strategies that have gained industrial acceptance in recent years. In addition to process monitoring, the related problem of monitoring the performance of the control system itself is also considered.

The general objectives of process monitoring are:

1. *Routine Monitoring*. Ensure that process variables are within specified limits.
2. *Detection and Diagnosis*. Detect abnormal process operation and diagnose the root cause.
3. *Preventive Monitoring*. Detect abnormal situations early enough that corrective action can be taken before the process is seriously upset.

Abnormal process operation can occur for a variety of reasons, including equipment problems (heat exchanger fouling), instrumentation malfunctions (sticking control valves, inaccurate sensors), and unusual disturbances (reduced catalyst activity, slowly drifting feed composition). Severe abnormal situations can have serious consequences, even forcing a plant shutdown. It has been estimated that improved handling of abnormal situations could result in savings of \$10 billion each year to the U.S. petrochemical industry (ASM, 2009). Thus, process monitoring and abnormal situation management are important activities.

The traditional approach for process monitoring is to compare measurements against specified limits. This *limit checking* technique is a standard feature of computer

control systems and is widely used to validate measurements of process variables such as flow rate, temperature, pressure, and liquid level. Process variables are measured quite frequently with sampling periods that typically are much smaller than the process settling time (see Chapter 17). However, for most industrial plants, many important *quality variables* cannot be measured on-line. Instead, samples of the product are taken on an infrequent basis (e.g., hourly or daily) and sent to the quality control laboratory for analysis. Due to the infrequent measurements, standard feedback control methods like PID control cannot be applied. Consequently, statistical process control techniques are implemented to ensure that the product quality meets the specifications.

The terms *statistical process control (SPC)* and *statistical quality control (SQC)* refer to a collection of statistically-based techniques that rely on *quality control charts* to monitor product quality. These terms tend to be used on an interchangeable basis. However, the term SPC is sometimes used to refer to a broader set of statistical techniques that are employed to improve process performance as well as product quality (MacGregor, 1988). In this chapter, we emphasize the classical SPC techniques that are based on quality control charts (also called control charts). The simplest control chart, a Shewhart chart, merely consists of measurements plotted vs. sample number, and control limits that indicate the upper and lower limits for normal process operation.

The major objective in SPC is to use process data and statistical techniques to determine whether the process operation is normal or abnormal. The SPC methodology is based on the fundamental assumption that normal process operation can be characterized by random variations about a mean value. If this situation exists, the process is said to be *in a state of statistical control* (or *in control*), and the control chart measurements tend to be normally distributed about the mean value. By contrast, frequent control chart violations would indicate abnormal process behavior or an *out-of-control* situation. Then, a search would be initiated to attempt to identify

the root cause of the abnormal behavior. The root cause is referred to as the *assignable cause* or the *special cause* in the SPC literature, while the normal process variability is referred to as *common cause* or *chance cause*. From an engineering perspective, SPC is more of a monitoring technique than a control technique because no automatic corrective action is taken after an abnormal situation is detected. A brief comparison of conventional feedback control and SPC is presented in Section 21.2.4. More detailed comparisons are available elsewhere (MacGregor, 1988; Box and Luceño, 1997).

The basic SPC concepts and control chart methodology were introduced by Shewhart (1931). The current widespread interest in SPC techniques began in the 1950s when they were successfully applied first in Japan and then in North America, Europe, and the rest of the world. Control chart methodologies are now widely used in discrete-parts manufacturing and in some sectors of the process industries, especially for the production of semiconductors, synthetic fibers, polymers, and specialty chemicals. SPC techniques are also widely used for product quality control and for monitoring control system performance (Shunta, 1995). The basic SPC methodology is described in introductory statistics texts (Montgomery and Runger, 2007) and books on SPC (Ryan 2000; Montgomery, 2009).

SPC techniques played a key role in the renewed industrial emphasis on product quality that is sometimes referred to as the *Quality Revolution*. During the 1980s, Deming (1986) had a major impact on industrial management in North America by convincing corporations that quality should be a top corporate priority. He argued that the failure of a company to produce quality products was largely a failure in management rather than a shortcoming of the plant equipment or employees. His success led to the establishment of many process and quality improvement programs, including the *Six Sigma* methodology that is considered in Section 21.3.

In this chapter, we first introduce traditional process monitoring techniques (Section 21.1) that are based on limit checking of measurements and process performance calculations. In Section 21.2, the theoretical basis of SPC monitoring techniques and the most widely used control charts are considered. We also introduce *process capability indices* and compare SPC with standard automatic feedback control. Traditional SPC monitoring techniques consider only a single measured variable at a time, a *univariate* approach. But when the measured variables are highly correlated, improved monitoring can be achieved by applying the *multivariate* techniques that are introduced in Section 21.4. In addition to monitoring process performance, it can be very beneficial to assess

control system performance. This topic is considered in Section 21.5.

Monitoring strategies have been proposed based on process models, neural networks, and expert systems (Davis et al., 2000; Chiang et al., 2001). However, these topics are beyond the scope of this book.

21.1 TRADITIONAL MONITORING TECHNIQUES

In this section, we consider two relatively simple but very effective process monitoring techniques: limit checking and performance calculations.

21.1.1 Limit Checking

Process measurements should be checked to ensure that they are between specified limits, a procedure referred to as *limit checking*. The most common types of measurement limits are (see Chapter 9):

1. High and low limits
2. High limit for the absolute value of the rate of change
3. Low limit for the sample variance

The limits are specified based on safety and environmental considerations, operating objectives, and equipment limitations. For example, the high limit on a reactor temperature could be set based on metallurgical limits or the onset of undesirable side reactions. The low limit for a slurry flow rate could be selected to avoid having solid material settle and plug the line. Sometimes a second set of limits serves as *warning limits*. For example, in a liquid storage system, when the level drops to 15% (the low limit), a low-priority alarm signal could be sent to the operator. But when the level decreases to 5% (the low-low limit), a high-priority alarm would be generated for this more serious situation. Similarly, in order to avoid having the tank overflow, a high limit of 85% and a high-high limit of 95% level could be specified. The high-high and low-low limits are also referred to as *action limits*.

In practice, there are physical limitations on how much a measurement can change between consecutive sampling instants. For example, we might conclude that a temperature in a process vessel cannot change by more than 2 °C from one sampling instant to the next, based on knowledge of the energy balance and the process dynamics. This *rate-of-change limit* can be used to detect an abnormal situation such as a noise spike or a sensor failure. (Noise-spike filters were considered in Chapter 17.)

A set of process measurements inevitably exhibits some variability, even for “steady-state operation.” This variability occurs as a result of measurement noise, turbulent flow near a sensor, and other process disturbances. However, if the amount of variability becomes

unusually low, it could indicate an abnormal situation such as a “dead sensor” or a sticking control valve. Consequently, it is common practice to monitor a measure of variability such as the variance or standard deviation of a set of measurements. For example, the variability of a set of n measurements can be characterized by the *sample standard deviation*, s , or the *sample variance*, s^2 ,

$$s^2 \triangleq \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (21-1)$$

where x_i denotes the i th measurement and \bar{x} is the sample mean:

$$\bar{x} \triangleq \frac{1}{n} \sum_{i=1}^n x_i \quad (21-2)$$

For a set of data, \bar{x} indicates the average value, while s and s^2 provide measures of the spread of the data. Either s or s^2 can be monitored to ensure that it is above a threshold that is specified based on process operating experience.

The flow rate data in Fig. 21.1 includes three noise spikes and a sensor failure. The rate of change limit would detect the noise spikes, while an abnormally low sample variance would identify the failed sensor. After a limit check violation occurs, an alarm signal can be sent to the plant operator in a number of different ways. A relatively minor alarm might merely be “logged” in a computer file. A more important alarm could be displayed as a flashing message on a computer terminal and require operator acknowledgment. A critical alarm could result in an audible sound or a flashing warning light in the control room. Other alarm options are available, as discussed in Chapter 9.

21.1.2 Performance Calculations

A variety of performance calculations can be made to determine whether the process and instrumentation are working properly. In particular, steady-state mass and energy balances are calculated using data that are

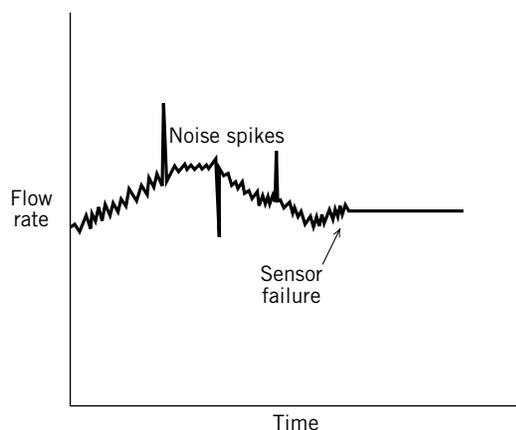


Figure 21.1 Flow rate measurement.

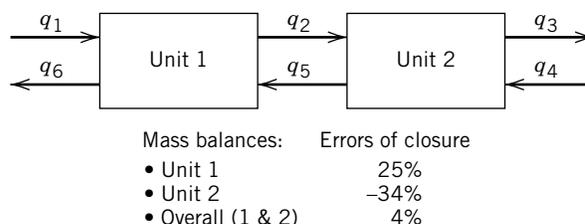


Figure 21.2 Countercurrent flow process.

averaged over a period of time (for example, one hour). The percent error of closure for a total mass balance can be defined as

$$\% \text{ error of closure} \triangleq \frac{\text{rate in} - \text{rate out}}{\text{rate in}} \times 100\% \quad (21-3)$$

A large error of closure may be caused by an equipment problem (e.g., a pipeline leak) or a sensor problem. *Data reconciliation* based on a statistical analysis of the errors of closure provides a systematic approach for deciding which measurements are suspect (Romagnoli and Sanchez, 2000).

Both redundant measurements and conservation equations can be used to good advantage. A process consisting of two units in a countercurrent flow configuration is shown in Fig. 21.2. Three steady-state mass balances can be written, one for each unit plus an overall balance around both units. Although the three balances are not independent, they provide useful information for monitoring purposes. Figure 21.2 indicates that the error of closure is small for the overall balance but large for each individual balance. This situation suggests that the flow rate sensor for one of the two interconnecting streams, q_2 or q_5 , may be faulty.

Process performance calculations also are very useful for diagnostic and monitoring purposes. For example, the thermal efficiency of a refrigeration unit or the selectivity of a chemical reactor could be calculated on a regular basis. A significant decrease from the normal value could indicate a process change or faulty measurement.

21.2 QUALITY CONTROL CHARTS

Industrial processes inevitably exhibit some variability in their manufactured products regardless of how well the processes are designed and operated. In statistical process control, an important distinction is made between normal (random) variability and abnormal (non-random) variability. Random variability is caused by the cumulative effects of a number of largely unavoidable phenomena such as electrical measurement noise, turbulence, and random fluctuations in feedstock or catalyst preparation. The random variability can be interpreted as a type of “background noise” for the manufacturing operation. Nonrandom variability can result from process changes (e.g., heat exchanger fouling, loss of catalyst

activity), faulty instrumentation, or human error. As mentioned earlier, the source of this abnormal variability is referred to as a *special cause* or an *assignable cause*.

21.2.1 Normal Distribution

Because the *normal distribution* plays a central role in SPC, we briefly review its important characteristics. The normal distribution is also known as the *Gaussian distribution*.

Suppose that a random variable x has a normal distribution with a mean μ and a variance σ^2 denoted by $N(\mu, \sigma^2)$. The probability that x has a value between two arbitrary constants, a and b , is given by:

$$P(a < x < b) = \int_a^b f(x)dx \quad (21-4)$$

where $P(\cdot)$ denotes the probability that x lies within the indicated range and $f(x)$ is the probability density function for the normal distribution:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right] \quad (21-5)$$

The following probability statements are valid for the normal distribution (Montgomery and Runger, 2007):

$$\begin{aligned} P(\mu - \sigma < x < \mu + \sigma) &= 0.6827 \\ P(\mu - 2\sigma < x < \mu + 2\sigma) &= 0.9545 \\ P(\mu - 3\sigma < x < \mu + 3\sigma) &= 0.9973 \end{aligned} \quad (21-6)$$

A graphical interpretation of these expressions is shown in Fig. 21.3 where each probability corresponds to an area under the $f(x)$ curve. Equation 21-6 and Fig. 21.3 demonstrate that if a random variable x is normally distributed, there is a very high probability (0.9973) that a measurement lies within 3σ of the mean μ . This important result provides the theoretical basis for widely used SPC techniques. Similar probability statements can be formulated based on statistical tables

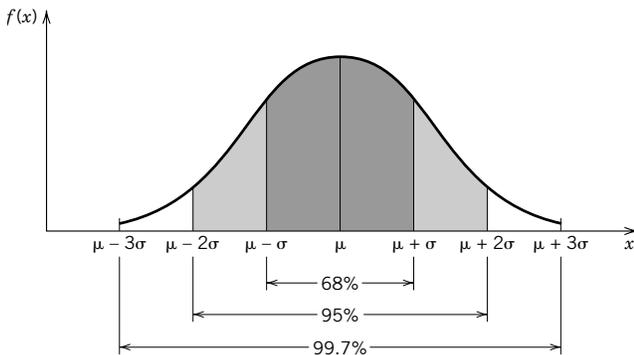


Figure 21.3 Probabilities associated with the normal distribution. From Montgomery and Runger (2007).

for the normal distribution. For the sake of generality, the tables are expressed in terms of the *standard normal distribution*, $N(0, 1)$, and the *standard normal variable*, $z \triangleq (x - \mu)/\sigma$.

It is important to distinguish between the theoretical mean μ and the sample mean \bar{x} . If measurements of a process variable are normally distributed, $N(\mu, \sigma^2)$, the sample mean is also normally distributed. Of course, for any particular sample, \bar{x} is not necessarily equal to μ .

21.2.2 The \bar{x} Control Chart

In statistical process control, Control Charts (or Quality Control Charts) are used to determine whether the process operation is normal or abnormal. The widely used \bar{x} control chart is introduced in the following example. This type of control chart is often referred to as a Shewhart Chart, in honor of the pioneering statistician, Walter Shewhart, who first developed it in the 1920s.

EXAMPLE 21.1

A manufacturing plant produces 10,000 plastic bottles per day. Because the product is inexpensive and the plant operation is normally satisfactory, it is not economically feasible to inspect every bottle. Instead, a sample of n bottles is randomly selected and inspected each day. These n items are called a *subgroup*, and n is referred to as the *subgroup size*. The inspection includes measuring the toughness x of each bottle in the subgroup and calculating the sample mean \bar{x} .

The \bar{x} control chart in Fig. 21.4 displays data for a 30-day period. The control chart has a *target* (T), an *upper control limit* (UCL), and a *lower control limit* (LCL). The target (or *centerline*) is the desired (or *expected*) value for \bar{x} , while the region between UCL and LCL defines the range of typical variability, as discussed below. If all of the \bar{x} data are within the control limits, the process operation is considered to be normal, or “in a state of control.” Data points outside the control limits are considered to be abnormal, indicating that the process operation is out of control. This situation occurs for the twenty-first sample. A single measurement located slightly beyond a control limit is not necessarily a cause for concern. But frequent or large chart violations should be investigated to determine a special cause.

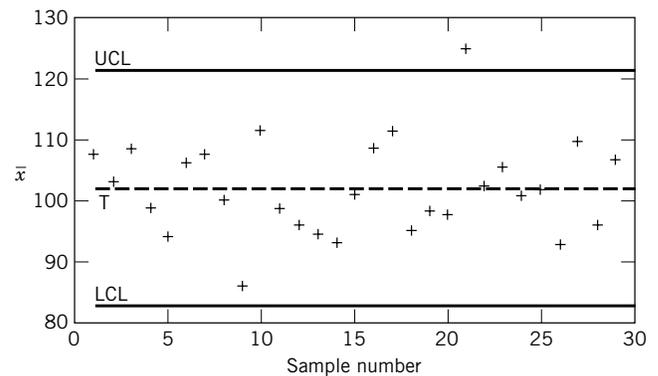


Figure 21.4 The \bar{x} control chart for Example 21.1.

The concept of a *rational subgroup* plays a key role in the development of quality control charts. The basic idea is that a subgroup should be specified so that it reflects typical process variability but not assignable causes. Thus, it is desirable to select a subgroup so that a special cause can be detected by a comparison of subgroups, but it will have little effect within a subgroup (Montgomery, 2009). For example, suppose that a small chemical plant includes six batch reactors and that a product quality measurement for each reactor is made every hour. If the monitoring objective is to determine whether overall production is satisfactory, then the individual reactor measurements could be pooled to provide a subgroup size of $n = 6$ and a sampling period of $\Delta t = 1$ h. On the other hand, if the objective is to monitor the performance of individual reactors, the product quality data for each reactor could be plotted on an hourly basis ($n = 1$) or averaged over an eight-hour shift ($n = 8$ and $\Delta t = 8$ h). When only a single measurement is made at each sampling instant, the subgroup size is $n = 1$ and the control chart is referred to as an *individuals chart*.

The first step in devising a control chart is to select a set of representative data for a period of time when the process operation is believed to be normal, rather than abnormal. Suppose that these test data consist of N subgroups that have been collected on a regular basis (for example, hourly or daily) and that each subgroup consists of n randomly selected items. Let x_{ij} denote the j th measurement in the i th subgroup. Then, the subgroup sample means can be calculated:

$$\bar{x}_i \triangleq \frac{1}{n} \sum_{j=1}^n x_{ij} \quad (i = 1, 2, \dots, N) \quad (21-7)$$

The *grand mean* $\bar{\bar{x}}$ is defined to be the average of the subgroup means:

$$\bar{\bar{x}} \triangleq \frac{1}{N} \sum_{i=1}^N \bar{x}_i \quad (21-8)$$

The general expressions for the control limits are

$$UCL \triangleq T + c\hat{\sigma}_{\bar{x}} \quad (21-9)$$

$$LCL \triangleq T - c\hat{\sigma}_{\bar{x}} \quad (21-10)$$

where $\hat{\sigma}_{\bar{x}}$ is an estimate of the standard deviation for \bar{x} and c is a positive integer; typically, $c = 3$. The choice of $c = 3$ and Eq. 21-6 imply that the measurements will lie within the control chart limits 99.73% of the time, for normal process operation. The target T is usually specified to be either $\bar{\bar{x}}$ or the desired value of \bar{x} .

The estimated standard deviation $\hat{\sigma}_{\bar{x}}$ can be calculated from the subgroups in the test data by two methods: (1) the standard deviation approach and (2) the range approach (Montgomery and Runger, 2007). By definition, the *range* R is the difference between the maximum and minimum values. Historically, the R approach has been emphasized, because R is easier to calculate than s , an

advantage for hand calculations. However, the standard deviation approach is now preferred because it uses all of the data, instead of only two points in each subgroup. It also has the advantage of being less sensitive to *outliers* (i.e., bad data points). However, for small values of n , the two approaches tend to produce similar control limits (Ryan, 2000). Consequently, we will only consider the standard deviation approach.

The average sample standard deviation \bar{s} for the N subgroups is

$$\bar{s} \triangleq \frac{1}{N} \sum_{i=1}^N s_i \quad (21-11)$$

where the standard deviation for the i th subgroup is

$$s_i \triangleq \sqrt{\frac{1}{n-1} \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2} \quad (21-12)$$

If the x data are normally distributed, then $\hat{\sigma}_{\bar{x}}$ is related to \bar{s} by

$$\hat{\sigma}_{\bar{x}} = \frac{1}{c_4 \sqrt{n}} \bar{s} \quad (21-13)$$

where c_4 is a constant that depends on n (Montgomery and Runger, 2007) and is tabulated in Table 21.1.

21.2.3 The s Control Chart

In addition to monitoring average process performance, it is also advantageous to monitor process variability. The variability within a subgroup can be characterized by its range, standard deviation, or sample variance. Control charts can be developed for all three statistics, but our discussion will be limited to the control chart for the standard deviation, the *s control chart*.

The centerline for the s chart is \bar{s} , which is the average standard deviation for the test set of data. The control limits are

$$UCL = B_4 \bar{s} \quad (21-14)$$

$$LCL = B_3 \bar{s} \quad (21-15)$$

Table 21.1 Control Chart Constants

Estimation of σ		s Chart	
n	c_4	B_3	B_4
2	0.7979	0	3.267
3	0.8862	0	2.568
4	0.9213	0	2.266
5	0.9400	0	2.089
6	0.9515	0.030	1.970
7	0.9594	0.118	1.882
8	0.9650	0.185	1.815
9	0.9693	0.239	1.761
10	0.9727	0.284	1.716
15	0.9823	0.428	1.572
20	0.9869	0.510	1.490
25	0.9896	0.565	1.435

Source: Adapted from Ryan (2000).

Constants B_3 and B_4 depend on the subgroup size n , as shown in Table 21.1.

The control chart limits for the \bar{x} and s charts in Eqs. 21-9 to 21-15 have been based on the assumption that the x data are normally distributed.

When individual measurements are plotted ($n = 1$), the standard deviation for the subgroup does not exist. In this situation, the *moving range* (MR) of two successive measurements can be employed to provide a measure of variability. The moving range is defined as the absolute value of the difference between successive measurements. Thus, for the k th sampling instant, $MR(k) = |x(k) - x(k-1)|$. The \bar{x} and s control charts are also applicable when the sample size n varies from one sample to the next.

Example 21.2 illustrates the construction of \bar{x} and s control charts.

EXAMPLE 21.2

In semiconductor processing, the photolithography process is used to transfer the circuit design to silicon wafers. In the first step of the process, a specified amount of a polymer solution, *photoresist*, is applied to a wafer as it spins at high speed on a turntable. The resulting photoresist thickness x is a key process variable. Thickness data for 25 subgroups are shown in Table 21.2. Each subgroup consists of three randomly

Table 21.2 Thickness Data (in Å) for Example 21.2

No.	x Data			\bar{x}	s
1	209.6	207.6	211.1	209.4	1.8
2	183.5	193.1	202.4	193.0	9.5
3	190.1	206.8	201.6	199.5	8.6
4	206.9	189.3	204.1	200.1	9.4
5	260.0	209.0	212.2	227.1	28.6
6	193.9	178.8	214.5	195.7	17.9
7	206.9	202.8	189.7	199.8	9.0
8	200.2	192.7	202.1	198.3	5.0
9	210.6	192.3	205.9	202.9	9.5
10	186.6	201.5	197.4	195.2	7.7
11	204.8	196.6	225.0	208.8	14.6
12	183.7	209.7	208.6	200.6	14.7
13	185.6	198.9	191.5	192.0	6.7
14	202.9	210.1	208.1	207.1	3.7
15	198.6	195.2	150.0	181.3	27.1
16	188.7	200.7	207.6	199.0	9.6
17	197.1	204.0	182.9	194.6	10.8
18	194.2	211.2	215.4	206.9	11.2
19	191.0	206.2	183.9	193.7	11.4
20	202.5	197.1	211.1	203.6	7.0
21	185.1	186.3	188.9	186.8	1.9
22	203.1	193.1	203.9	200.0	6.0
23	179.7	203.3	209.7	197.6	15.8
24	205.3	190.0	208.2	201.2	9.8
25	203.4	202.9	200.4	202.2	1.61

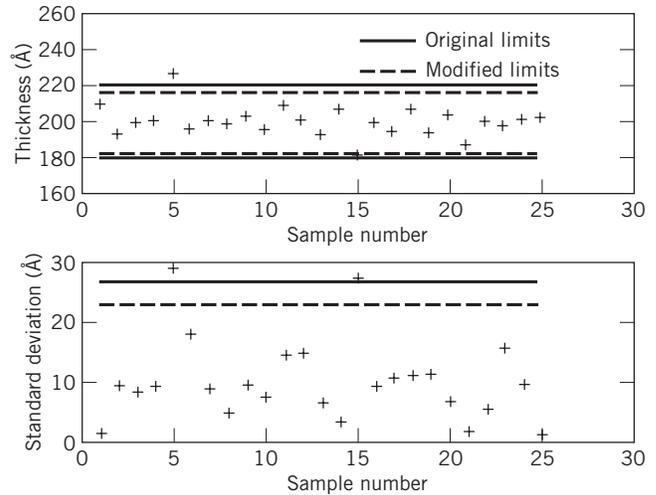


Figure 21.5 The \bar{x} and s control charts for Example 21.2.

selected wafers. Construct \bar{x} and s control charts for these test data and critically evaluate the results.

SOLUTION

The following sample statistics can be calculated from the data in Table 21.2: $\bar{\bar{x}} = 199.8 \text{ Å}$, $\bar{\bar{s}} = 10.4 \text{ Å}$. For $n = 3$ the required constants from Table 21.1 are $c_4 = 0.8862$, $B_3 = 0$, and $B_4 = 2.568$. Then the \bar{x} and s control limits can be calculated from Eqs. 21-9 to 21-15. The traditional value of $c = 3$ is selected for Eqs. 21-9 and 21-10. The resulting control limits are labeled as the “original limits” in Fig. 21.5.

Figure 21.5 indicates that sample #5 lies beyond the UCL for both the \bar{x} and s control charts, while sample #15 is very close to a control limit on each chart. Thus, the question arises whether these two samples are “outliers” that should be omitted from the analysis. Table 21.2 indicates that sample #5 includes a very large value (260.0), while sample #15 includes a very small value (150.0). However, unusually large or small numerical values by themselves do not justify discarding samples; further investigation is required.

Suppose that a more detailed evaluation has discovered a specific reason as to why measurements #5 and #15 should be discarded (e.g., faulty sensor, data misreported, etc.). In this situation, these two samples should be removed and the control limits should be recalculated based on the remaining 23 samples. These modified control limits are tabulated below as well as in Fig. 21.5.

	Original Limits	Modified Limits (omit samples #5 and #15)
\bar{x} Chart Control Limits		
UCL	220.1	216.7
LCL	179.6	182.2
s Chart Control Limits		
UCL	26.6	22.7
LCL	0	0

21.2.4 Theoretical Basis for Quality Control Charts

The traditional SPC methodology is based on the assumption that the natural variability for “in control” conditions can be characterized by random variations around a constant average value,

$$x(k) = x^* + e(k) \quad (21-16)$$

where $x(k)$ is the measurement at time k , x^* is the true (but unknown) value, and $e(k)$ is an additive random error. Traditional control charts are based on the following assumptions:

1. Each additive error, $\{e(k), k = 1, 2, \dots\}$, is a zero-mean, random variable that has the same normal distribution, $N(0, \sigma^2)$.
2. The additive errors are statistically independent and thus uncorrelated. Consequently, $e(k)$ does not depend on $e(j)$ for $j \neq k$.
3. The true value x^* is constant.
4. The subgroup size n is the same for all of the subgroups.

The second assumption is referred to as *independent and identically distributed (IID)*.

Consider an ideal individuals control chart for x with x^* as its target and “ 3σ control limits”:

$$UCL \triangleq x^* + 3\sigma \quad (21-17)$$

$$LCL \triangleq x^* - 3\sigma \quad (21-18)$$

These control limits are a special case of Eqs. 21-9 and 21-10 for the idealized situation where σ is known, $c = 3$, and the subgroup size is $n = 1$. The typical choice of $c = 3$ can be justified as follows. Because x is $N(0, \sigma^2)$, the probability p that a measurement lies outside the 3σ control limits can be calculated from Eq. 21-6: $p = 1 - 0.9973 = 0.0027$. Thus on average, approximately three out of every 1,000 measurements will be outside of the 3σ limits. The average number of samples before a chart violation occurs is referred to as the *average run length (ARL)*. For the normal (“in control”) process operation,

$$ARL \triangleq \frac{1}{p} = \frac{1}{0.0027} = 370 \quad (21-19)$$

Thus, a Shewhart chart with 3σ control limits will have an average of one control chart violation every 370 samples, even when the process is *in a state of control*.

This theoretical analysis justifies the use of 3σ limits for \bar{x} and other control charts. However, other values of c are sometimes used. For example, 2σ warning limits can be displayed on the control chart in addition to the 3σ control limits. Although the 2σ warning limits provide an early indication of a process change, they have a very low average run length value of $ARL = 22$. In general, larger values of c result in wider chart limits and larger ARL values. Wider chart limits mean that

process changes will not be detected as quickly as they would be for smaller c values. Thus, the choice of c involves a classical engineering compromise between early detection of process changes (low value of c) and reducing the frequency of false alarms (high value of c).

Standard SPC techniques are based on the four assumptions listed above. However, because these assumptions are not always valid for industrial processes, standard techniques can give misleading results. In particular, the implications of violating the normally distributed and IID assumptions have received considerable theoretical analysis (Ryan, 2000). Although modified SPC techniques have been developed for these nonideal situations, commercial SPC software is usually based on these assumptions.

Industrial plant measurements are not normally distributed. However, for large subgroup sizes ($n > 25$), \bar{x} is approximately normally distributed even if x is not, according to the famous *Central Limit Theorem* of statistics (Montgomery and Runger, 2007). Fortunately, modest deviations from “normality” can be tolerated. In addition, the standard SPC techniques can be modified so that they are applicable to certain classes of nonnormal data (Jacobs, 1990).

In industrial applications, the control chart data are often *serially correlated*, because the current measurement is related to previous measurements. For example, the flow rate data in Fig. 21.1 are serially correlated. Standard control charts such as the \bar{x} and s charts can provide misleading results if the data are serially correlated. But if the degree of correlation is known, the control limits can be adjusted accordingly (Montgomery, 2009). Serially correlated data also can be modeled using time-series analysis, as described in Section 17.6.

21.2.5 Pattern Tests and the Western Electric Rules

We have considered how abnormal process behavior can be detected by comparing individual measurements with the \bar{x} and s control chart limits. However, the pattern of measurements can also provide useful information. For example, if 10 consecutive measurements are all increasing, then it is very unlikely that the process is *in a state of control*.

A wide variety of pattern tests (also called *zone rules*) can be developed based on the IID and normal distribution assumptions and the properties of the normal distribution. For example, the following excerpts from the *Western Electric Rules* (Western Electric Company, 1956; Montgomery and Runger, 2007) indicate that the process is *out of control* if one or more of the following conditions occur:

1. One data point is outside the 3σ control limits.
2. Two out of three consecutive data points are beyond a 2σ limit.

3. Four out of five consecutive data points are beyond a 1σ limit and on one side of the centerline.
4. Eight consecutive points are on one side of the centerline.

Note that the first condition corresponds to the familiar Shewhart chart limits of Eqs. 21-9 and 21-10 with $c = 3$. Additional pattern tests are concerned with other types of nonrandom behavior (Montgomery, 2009). Pattern tests can be used to augment Shewhart charts. This combination enables out-of-control behavior to be detected earlier, but the false alarm rate is higher than that for a Shewhart chart alone.

21.2.6 CUSUM and EWMA Control Charts

Although Shewhart charts with 3σ limits can quickly detect large process changes, they are ineffective for small, sustained process changes (for example, changes in μ smaller than 1.5σ). Two alternative control charts have been developed to detect small changes: the CUSUM and EWMA control charts. They also can detect large process changes (for example, 3σ shifts), but detection is usually somewhat slower than for Shewhart charts. Because the CUSUM and EWMA control charts can effectively detect both large and small process shifts, they provide viable alternatives to the widely used Shewhart charts. Consequently, they will now be considered. The *cumulative sum (CUSUM)* is defined to be a running summation of the deviations of the plotted variable from its target. If the sample mean is plotted, the cumulative sum, $C(k)$, is

$$C(k) = \sum_{j=1}^k (\bar{x}(j) - T) \quad (21-20)$$

where T is the target for \bar{x} . During normal process operation, $C(k)$ fluctuates around zero. But if a process change causes a small shift in \bar{x} , $C(k)$ will drift either upward or downward.

The CUSUM control chart was originally developed using a graphical approach based on *V-masks* (Montgomery, 2009). However, for computer calculations, it is more convenient to use an equivalent algebraic version that consists of two recursive equations,

$$C^+(k) = \max[0, \bar{x}(k) - (T + K) + C^+(k - 1)] \quad (21-21)$$

$$C^-(k) = \max[0, (T - K) - \bar{x}(k) + C^-(k - 1)] \quad (21-22)$$

where C^+ and C^- denote the sums for the high and low directions and K is a constant, the *slack parameter*. The CUSUM calculations are initialized by setting $C^+(0) = C^-(0) = 0$. A deviation from the target that is larger than K increases either C^+ or C^- . A control limit violation occurs when either C^+ or C^- exceeds a specified control limit (or *threshold*), H . After a limit violation occurs, that sum is reset to zero or to a specified value.

The selection of the threshold H can be based on considerations of average run length. Suppose that we want

Table 21.3 Average Run Lengths for CUSUM Control Charts

Shift from Target (in multiples of $\hat{\sigma}_{\bar{x}}$)	ARL for $H = 4\hat{\sigma}_{\bar{x}}$	ARL for $H = 5\hat{\sigma}_{\bar{x}}$
0	168.0	465.0
0.25	74.2	139.0
0.50	26.6	38.0
0.75	13.3	17.0
1.00	8.38	10.4
2.00	3.34	4.01
3.00	2.19	2.57

Source: Adapted from Ryan (2000).

to detect whether the sample mean \bar{x} has shifted from the target by a small amount, δ . The slack parameter K is usually specified as $K = 0.5 \delta$. For the ideal situation where the normally distributed and IID assumptions are valid, ARL values have been tabulated for specified values of δ , K , and H (Ryan, 2000; Montgomery, 2009).

Table 21.3 summarizes ARL values for two values of H and different values of δ . (The values of δ are usually expressed as multiples of $\hat{\sigma}_{\bar{x}}$.) The ARL values indicate the average number of samples before a change of δ is detected. Thus, the ARL values for $\delta = 0$ indicate the average time between “false alarms,” that is, the average time between successive CUSUM alarms when no shift in \bar{x} has occurred. Ideally, we would like the ARL value to be very large for $\delta = 0$, and small for $\delta \neq 0$. Table 21.3 shows that as the magnitude of the shift δ increases, ARL decreases, and thus the CUSUM control chart detects the change faster. Increasing the value of H from 4σ to 5σ increases all of the ARL values and thus provides a more conservative approach.

CUSUM control charts also are constructed for measures of variability such as the range or standard deviation (Ryan, 2000; Montgomery, 2009).

EWMA Control Chart

Information about past measurements can also be included in the control chart calculations by exponentially weighting the data. This strategy provides the basis for the *exponentially weighted moving-average (EWMA) control chart*. Let \bar{x} denote the sample mean of the measured variable and z denote the EWMA of \bar{x} . A recursive equation is used to calculate $z(k)$,

$$z(k) = \lambda \bar{x}(k) + (1 - \lambda)z(k - 1) \quad (21-23)$$

where λ is a constant, $0 \leq \lambda \leq 1$. Note that Eq. 21-23 has the same form as the first-order (or exponential) filter that was introduced in Chapter 17. The EWMA control chart consists of a plot of $z(k)$ vs. k , as well as a target and upper and lower control limits. Note that the EWMA control chart reduces to the Shewhart chart for $\lambda = 1$. The EWMA calculations are initialized by setting $z(0) = T$.

If the \bar{x} measurements satisfy the IID condition, the EWMA control limits can be derived. The theoretical 3σ limits are given by

$$T \pm 3\hat{\sigma}_{\bar{x}} \sqrt{\frac{\lambda}{2-\lambda}} \quad (21-24)$$

where $\hat{\sigma}_{\bar{x}}$ is determined from a set of test data taken when the process is *in a state of control* (Montgomery, 2009). The target T is selected to be either the desired value of \bar{x} or the grand mean for the test data, $\bar{\bar{x}}$. Time-varying control limits can also be derived that provide narrower limits for the first few samples, for applications where early detection is important (Montgomery, 2009; Ryan, 2000). Tables of ARL values have been developed for the EWMA method, similar to Table 21.3 for the CUSUM method (Ryan, 2000).

The EWMA performance can be adjusted by specifying λ . For example, $\lambda = 0.25$ is a reasonable choice, because it results in an ARL of 493 for no mean shift ($\delta = 0$) and an ARL of 11 for a mean shift of $\sigma_{\bar{x}}$ ($\delta = 1$). EWMA control charts can also be constructed for measures of variability such as the range and standard deviation.

EXAMPLE 21.3

In order to compare Shewhart, CUSUM, and EWMA control charts, consider simulated data for the tensile strength of a phenolic resin. It is assumed that the tensile strength x is normally distributed with a mean of $\mu = 70$ MPa and a standard deviation of $\sigma = 3$ MPa. A single measurement is available at each sampling instant. A constant ($\delta = 0.5\sigma = 1.5$) was added to $x(k)$ for $k \geq 10$ in order to evaluate each chart's ability to detect a small process shift. The CUSUM chart was designed using $K = 0.5\sigma$ and $H = 5\sigma$, while the EWMA parameter was specified as $\lambda = 0.25$.

The relative performance of the Shewhart, CUSUM, and EWMA control charts is compared in Fig. 21.6. The Shewhart

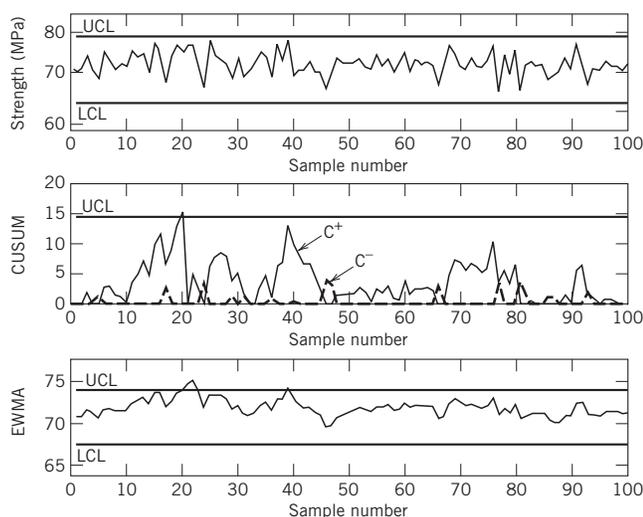


Figure 21.6 Comparison of Shewhart (top), CUSUM (middle), and EWMA (bottom) control charts for Example 21.3.

chart fails to detect the 0.5σ shift in x . However, both the CUSUM and EWMA charts quickly detect this change, because limit violations occur about 10 samples after the shift occurs (at $k = 20$ and $k = 21$, respectively). The mean shift can also be detected by applying the Western Electric Rules in the previous section.

21.3 EXTENSIONS OF STATISTICAL PROCESS CONTROL

Now that the basic quality control charts have been presented, we consider several other important topics in statistical process control.

21.3.1 Process Capability Indices

Process capability indices (or *process capability ratios*) provide a measure of whether an “in control” process is meeting its product specifications. Suppose that a quality variable x must have a volume between an *upper specification limit (USL)* and a *lower specification limit (LSL)* in order for product to satisfy customer requirements. The C_p *capability index* is defined as

$$C_p \triangleq \frac{USL - LSL}{6\sigma} \quad (21-25)$$

where σ is the standard deviation of x . Suppose that $C_p = 1$ and x is normally distributed. Based on Eq. 21-6, we would expect that 99.73% of the measurements satisfy the specification limits. If $C_p > 1$, the product specifications are satisfied; for $C_p < 1$, they are not.

A second capability index C_{pk} is based on average process performance (\bar{x}), as well as process variability (σ). It is defined as

$$C_{pk} \triangleq \frac{\min[\bar{x} - LSL, USL - \bar{x}]}{3\sigma} \quad (21-26)$$

Although both C_p and C_{pk} are used, we consider C_{pk} to be superior to C_p for the following reason. If $\bar{x} = T$, the process is said to be “centered” and $C_{pk} = C_p$. But for $\bar{x} \neq T$, C_p does not change, even though the process performance is worse, while C_{pk} decreases. For this reason, C_{pk} is preferred.

If the standard deviation σ is not known, it is replaced by an estimate $\hat{\sigma}$ in Eqs. 21-25 and 21-26. For situations where there is only a single specification limit, either USL or LSL, the definitions of C_p and C_{pk} can be modified accordingly (Ryan, 2000).

In practical applications, a common objective is to have a capability index of 2.0, while a value greater than 1.5 is considered to be acceptable (Shunta, 1995). If the C_{pk} value is too low, it can be improved by making a change that either reduces process variability or causes \bar{x} to move closer to the target. These improvements can be achieved in a number of ways, including better process control, better process maintenance,

reduced variability in raw materials, improved operator training, and changes in process operating conditions.

Three important points should be noted concerning the C_p and C_{pk} capability indices:

1. The data used in the calculations do *not* have to be normally distributed.
2. The specification limits, USL and LSL, and the control limits, UCL and LCL, are *not* related. The specification limits denote the desired process performance, while the control limits represent actual performance during normal operation when the process is *in control*.
3. The numerical values of the C_p and C_{pk} capability indices in (21-25) and (21-26) are only meaningful when the process is *in a state of control*. However, other *process performance indices* are available to characterize process performance when the process is *not in a state of control*. They can be used to evaluate the incentives for improved process control (Shunta, 1995).

EXAMPLE 21.4

Calculate the average values of the C_p and C_{pk} capability indices for the photolithography thickness data in Example 21.2. Omit the two outliers (samples #5 and #15), and assume that the upper and lower specification limits for the photoresist thickness are $USL = 235 \text{ \AA}$ and $LSL = 185 \text{ \AA}$.

SOLUTION

After samples #5 and #15 are omitted, the grand mean is $\bar{x} = 199 \text{ \AA}$, and the standard deviation of \bar{x} (estimated from Eq. 21-13 with $c_4 = 0.8862$) is

$$\hat{\sigma}_{\bar{x}} = \frac{\bar{s}}{c_4 \sqrt{n}} = \frac{8.83}{0.8862 \sqrt{3}} = 5.75 \text{ \AA}$$

From Eqs. 21-25 and 21-26,

$$C_p = \frac{235 - 185}{6(5.75)} = 1.45$$

$$C_{pk} = \frac{\min[199.5 - 185, 235 - 199.5]}{3(5.75)} = 0.84$$

Note that C_{pk} is much smaller than the C_p , because \bar{x} is closer to the LSL than the USL.

21.3.2 Six Sigma Approach

Product quality specifications continue to become more stringent as a result of market demands and intense worldwide competition. Meeting quality requirements is especially difficult for products that consist of a very large number of components and for manufacturing processes that consist of hundreds of individual steps. For example, the production of a microelectronics device typically requires 100 to 300 batch processing steps. Suppose that there are 200 steps, and that

each one must meet a quality specification in order for the final product to function properly. If each step, is independent of the others and has a 99% success rate, the overall yield of satisfactory product is $(0.99)^{200} = 0.134$, or only 13.4%. This low yield is clearly unsatisfactory. Similarly, even when a processing step meets 3σ specifications (99.73% success rate), it will still result in an average of 2,700 “defects” for every million produced. Furthermore, the overall yield for this 200-step process is still only 58.2%.

These examples demonstrate that for complicated products or processes, 3σ quality is no longer adequate, and there is no place for failure. These considerations and economic pressures have motivated the development of the *six sigma approach* (Pande et al., 2000). The statistical motivation for this approach is based on the properties of the normal distribution. Suppose that a product quality variable x is normally distributed, $N(\mu, \sigma^2)$. As indicated on the left portion of Fig. 21.7, if the product specifications are $\mu \pm 6\sigma$, the product will meet the specifications 99.99998% of the time. Thus, on average, there will only be two defective products for every billion produced. Now suppose that the process operation changes

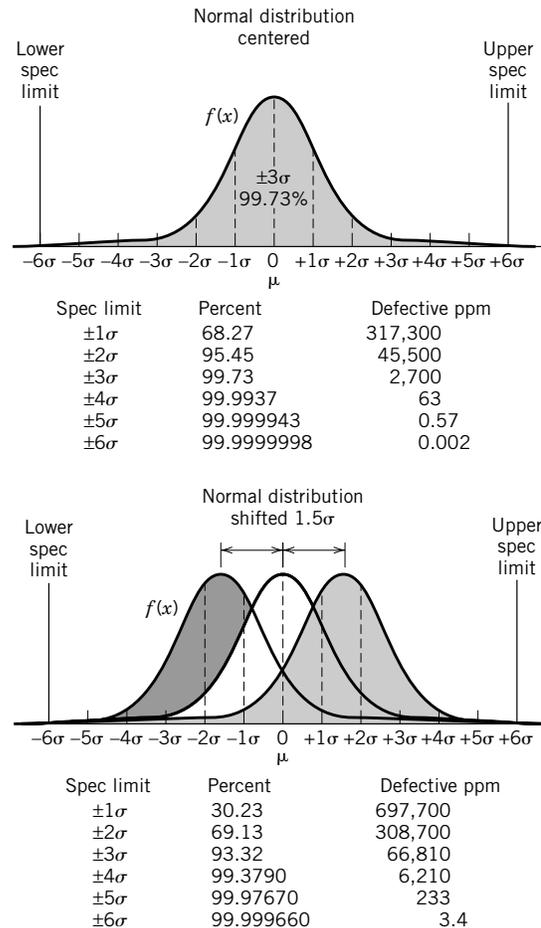


Figure 21.7 The Six Sigma Concept (Montgomery and Runger, 2007). Top: No shift in the mean. Bottom: 1.5σ shift.

so that the mean value is shifted from $\bar{x} = \mu$ to either $\bar{x} = \mu + 1.5\sigma$ or $\bar{x} = \mu - 1.5\sigma$, as shown on the right side of Fig. 21.7. Then the product specifications will still be satisfied 99.99966% of the time, which corresponds to 3.4 defective products per million produced.

In summary, if the variability of a manufacturing operation is so small that the product specification limits are equal to $\mu \pm 6\sigma$, then the limits can be satisfied even if the mean value of x shifts by as much as 1.5σ . This very desirable situation of near perfect product quality is referred to as *six sigma quality*.

The six sigma approach was pioneered by the Motorola and General Electric companies in the early 1980s as a strategy for achieving both six sigma quality and continuous improvement. Since then, other large corporations have adopted companywide programs that apply the six sigma approach to all of their business operations, both manufacturing and nonmanufacturing. Thus, although the six sigma approach is “data-driven” and based on statistical techniques, it has evolved into a broader management philosophy that has been implemented successfully by many large corporations. Six sigma programs have also had a significant financial impact. Large corporations have reported savings of billions of dollars that were attributed to successful six sigma programs.

In summary, the six sigma approach based on statistical monitoring techniques has had a major impact on both manufacturing and business practice during the past two decades. It is based on SPC concepts but has evolved into a much broader management philosophy and corporatewide activity. Improved process control can play a key role in a six sigma project by reducing the variability in controlled variables that have a significant economic impact.

21.3.3 Comparison of Statistical Process Control and Automatic Process Control

Statistical process control and automatic process control (APC) are complementary techniques that were developed for different types of problems. As indicated in earlier chapters, APC takes corrective action when a controlled variable deviates from the set point. The corrective action tends to change at each sampling instant. Thus, for APC there is an implicit assumption that the cost of making a corrective action is not significant. APC is widely used in the process industries, because no information is required about the sources and types of process disturbances. APC is most effective when the measurement sampling period is relatively short compared to the process settling time, and when the process disturbances tend to be deterministic (that is, when they have a sustained nature such as a step or ramp disturbance).

In statistical process control, the objective is to decide whether the process is behaving normally, and to identify

a special cause when it is not. In contrast to APC, no corrective action is taken when the measurements are within the control chart limits. This philosophy is appropriate when there is a significant cost associated with taking a corrective action, such as when shutting down a process unit or taking an instrument out of service for maintenance. From an engineering perspective, SPC is viewed as a *monitoring*, rather than a *control*, strategy. It is very effective when the normal process operation can be characterized by random fluctuations around a mean value. SPC is an appropriate choice for monitoring problems where the sampling period is long compared to the process settling time and the process disturbances tend to be random rather than deterministic. SPC has been widely used for quality control in both discrete-parts manufacturing and the process industries.

In summary, SPC and APC should be regarded as complementary rather than competitive techniques. They were developed for different types of situations and have been successfully used in the process industries. Furthermore, a combination of the two methods can be very effective. For example, in model-based control such as model predictive control (Chapter 20), APC can be used for feedback control, while SPC is used to monitor the model residuals, the differences between the model predictions and the actual values.

21.4 MULTIVARIATE STATISTICAL TECHNIQUES

In Chapters 12 and 16, we have emphasized that many important control problems are multivariable in nature because more than one process variable must be controlled and more than one variable can be manipulated. Similarly, for common SPC monitoring problems, two or more quality variables are important, and they can be highly correlated. For example, 10 or more quality variables are typically measured for synthetic fibers (MacGregor, 1996). For these situations, multivariable SPC techniques can offer significant advantages over the single-variable methods discussed in Section 21.2. In the statistics literature, these techniques are referred to as *multivariate methods*, while the standard Shewhart and CUSUM control charts are examples of *univariate methods*. The advantage of a multivariate monitoring approach is illustrated in Example 21.5.

EXAMPLE 21.5

The effluent stream from a wastewater treatment process is monitored to make sure that two process variables, the biological oxidation demand (BOD) and the solids content, meet specifications. Representative data are shown in Table 21.4. Shewhart charts for the sample means are shown in parts (a) and (b) of Fig. 21.8. These univariate control charts indicate that the process appears to be in-control because no chart

Table 21.4 Wastewater Treatment Data

Sample Number	BOD (mg/L)	Solids (mg/L)
1	17.7	1380
2	23.6	1458
3	13.2	1322
4	25.2	1448
5	13.1	1334
6	27.8	1485
7	29.8	1503
8	9.0	1540
9	14.3	1341
10	26.0	1448
11	23.2	1426
12	22.8	1417
13	20.4	1384
14	17.5	1380
15	18.4	1396
16	16.8	1345
17	13.8	1349
18	19.4	1398
19	24.7	1426
20	16.8	1361
21	14.9	1347
22	27.6	1476
23	26.1	1454
24	20.0	1393
25	22.9	1427
26	22.4	1431
27	19.6	1405
28	31.5	1521
29	19.9	1409
30	20.3	1392

violations occur for either variable. However, the bivariate control chart in Fig. 21.8c indicates that the two variables are highly correlated, because the solids content tends to be large when the BOD is large, and vice versa. When the two variables are considered together, their joint confidence limit (e.g., at the 99% confidence level) is an ellipse, as shown in Fig. 21.8c.¹ Sample #8 lies well beyond the 99% limit, indicating an out-of-control condition. By contrast, this sample lies within the Shewhart control chart limits for both individual variables.

This example has demonstrated that univariate SPC techniques such as Shewhart charts can fail to detect abnormal process behavior when the process variables are highly correlated. By contrast, the abnormal situation was readily apparent from the multivariate analysis.

Figure 21.9 provides a general comparison of univariate and multivariate SPC techniques (Alt et al., 1998). When two variables, x_1 and x_2 , are monitored individually, the two sets of control limits define a rectangular region, as shown in Fig. 21.9. In analogy with Example 21.5, the multivariate control limits define the dark, ellipsoidal region that represents *in-control* behavior. Figure 21.9 demonstrates that the application of univariate SPC techniques to correlated multivariate data can result in two types of misclassification: false alarms and out-of-control conditions that are not detected. The latter type of misclassification occurred at sample #8 for the two Shewhart charts in Fig. 21.8.

In the next section, we consider some well-known multivariate monitoring techniques.

21.4.1 Hotelling's T^2 Statistic

Suppose that it is desired to use SPC techniques to monitor p variables, which are correlated and normally distributed. Let \mathbf{x} denote the column vector of these p variables, $\mathbf{x} = \text{col}[x_1, x_2, \dots, x_p]$. At each sampling instant, a subgroup of n measurements is made for each variable. The subgroup sample means for the k th sampling instant can be expressed as a column vector: $\bar{\mathbf{x}}(k) = \text{col}[\bar{x}_1(k), \bar{x}_2(k), \dots, \bar{x}_p(k)]$. Multivariate control charts are traditionally based on *Hotelling's T^2 statistic* (Montgomery, 2009).

$$T^2(k) \triangleq n[\bar{\mathbf{x}}(k) - \bar{\bar{\mathbf{x}}}]^T \mathbf{S}^{-1} [\bar{\mathbf{x}}(k) - \bar{\bar{\mathbf{x}}}] \quad (21-27)$$

where $T^2(k)$ denotes the value of the T^2 statistic at the k th sampling instant. The vector of grand means $\bar{\bar{\mathbf{x}}}$

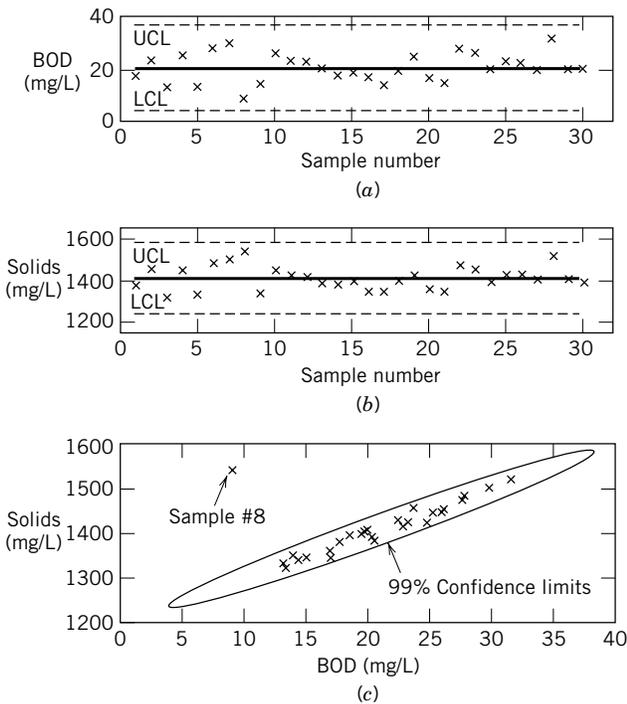


Figure 21.8 Confidence regions for Example 21.5. Univariate in (a) and (b), bivariate in (c).

¹If two random variables are correlated and normally distributed, the confidence limit is in the form of an ellipse and can be calculated from the well-known F distribution (Montgomery and Runger, 2007).

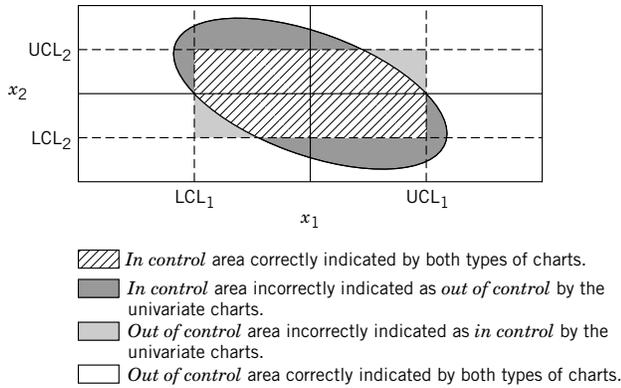


Figure 21.9 Univariate and bivariate confidence regions for two random variables, x_1 and x_2 (modified from Alt et al., 1998).

and the covariance matrix \mathbf{S} are calculated for a test set of data for *in-control* conditions. By definition S_{ij} , the (i, j) -element of matrix \mathbf{S} , is the sample covariance of x_i and x_j :

$$S_{ij} \triangleq \frac{1}{N} \sum_{k=1}^N [\bar{x}_i(k) - \bar{\bar{x}}_i] [\bar{x}_j(k) - \bar{\bar{x}}_j] \quad (21-28)$$

In Eq. 21-28 N is the number of subgroups, and $\bar{\bar{x}}_i$ denotes the mean for \bar{x}_i .

Note that T^2 is a scalar, even though the other quantities in Eq. 21-27 are vectors and matrices. The inverse of the sample covariance matrix, \mathbf{S}^{-1} , scales the p variables and accounts for correlation among them.

A multivariate process is considered to be out-of-control at the k th sampling instant if $T^2(k)$ exceeds an upper control limit (UCL). (There is no target or lower control limit.) The UCL values are tabulated in statistics books and depend on the number of variables p and the subgroup size n . The T^2 control chart consists of a plot of $T^2(k)$ vs. k and an UCL. Thus, the T^2 control chart is the multivariate generalization of the \bar{x} chart introduced in Section 21.2.2. Multivariate generalizations of the CUSUM and EWMA charts are also available (Montgomery, 2009).

EXAMPLE 21.6

Construct a T^2 control chart for the wastewater treatment problem of Example 21.5. The 99% control chart limit is $T^2 = 11.63$. Is the number of T^2 control chart violations consistent with the results of Example 21.5?

SOLUTION

The T^2 control chart is shown in Fig. 21.10. All of the T^2 values lie below the 99% confidence limit except for sample #8. This result is consistent with the bivariate control chart in Fig. 21.8c.

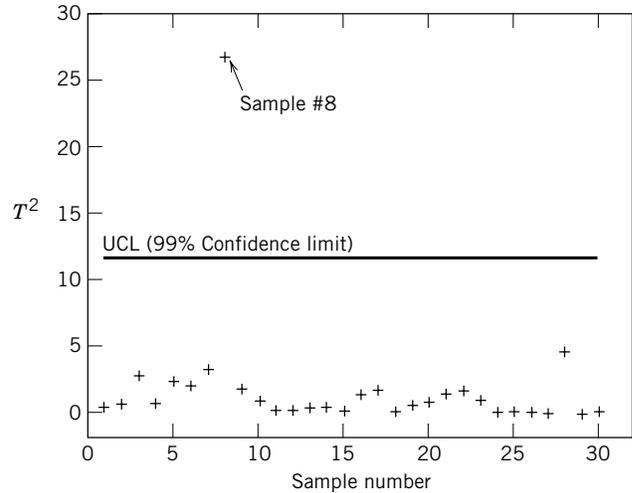


Figure 21.10 T^2 control chart for Example 21.5.

21.4.2 Principal Component Analysis and Partial Least Squares

Multivariate monitoring based on Hotelling’s T^2 statistic can be effective if the data are not highly correlated and the number of variables p is not large (for example, $p < 10$). For highly correlated data, the \mathbf{S} matrix is poorly conditioned and the T^2 approach becomes problematic. Fortunately, alternative multivariate monitoring techniques have been developed that are very effective for monitoring problems with large numbers of variables and highly correlated data. The *Principal Component Analysis (PCA)* and *Partial Least Squares (PLS)* methods have received the most attention in the process control community. Both techniques can be used to monitor process variables (e.g., temperature, level, pressure, and flow measurements) as well as product quality variables. These methods can provide useful diagnostic information after a chart violation has been detected. Although the PCA and PLS methods are beyond the scope of this book, excellent books (Jackson, 1991; Piovoso and Khosanovich, 1996; Montgomery, 2009), survey articles (Kourti, 2002) and a special issue of a journal (Piovoso and Hoo, 2002) are available.

21.5 CONTROL PERFORMANCE MONITORING

In order to achieve the desired process operation, the control system must function properly. As indicated in Chapter 11, industrial surveys have reported that many control loops perform poorly and even increase variability in comparison with manual control. Contributing factors include poor controller tuning and control valves that are incorrectly sized or tend to

Table 21.5 Basic Data for Control Loop Monitoring

-
- Service factors (time in use/total time period)
 - Mean and standard deviation for the control error (set point – measurement)
 - Mean and standard deviation for the controller output
 - Alarm summaries
 - Operator logbooks and maintenance records
-

stick due to excessive frictional forces. In large processing plants, each plant operator is typically responsible for 200 to 1,000 loops. Thus, there are strong incentives for automated *control* (or *controller*) *performance monitoring* (CPM). The overall objectives of CPM are (1) to determine whether the control system is performing in a satisfactory manner and (2) to diagnose the cause of any unsatisfactory performance.

21.5.1 Basic Information for Control Performance Monitoring

In order to monitor the performance of a single standard PI or PID control loop, the basic information in Table 21.5 should be available.

Service factors should be calculated for key components of the control loop such as the sensor and final control element. Low service factors and/or frequent maintenance suggest chronic problems that require attention. The fraction of time that the controller is in the automatic mode is a key metric. A low value indicates that the loop is frequently in the manual mode and thus requires attention. Service factors for computer hardware and software should also be recorded.

Simple statistical measures such as the sample mean and standard deviation can indicate whether the controlled variable is achieving its target and how much control effort is required. An unusually small standard deviation for a measurement could result from a faulty sensor with a constant output signal, as noted in Section 21.1. By contrast, an unusually large standard deviation could be caused by equipment degradation or even failure, for example, inadequate mixing caused by a faulty vessel agitator.

A high alarm rate can be indicative of poor control system performance (see Section 9.2). Operator logbooks and maintenance records are valuable sources of information, especially if this information has been captured in a computer database.

SUMMARY

Process monitoring is essential to ensure that plants operate safely and economically while meeting environmental standards. In recent years, control system performance

21.5.2 Control Performance Monitoring Techniques

Chapters 5 and 11 introduced traditional control loop performance criteria such as rise time, settling time, overshoot, offset, degree of oscillation, and integral error criteria. CPM methods have been developed based on these and other criteria, and commercial CPM software is available. A comprehensive review of CPM techniques and industrial applications has been reported by Jelali (2006).

If a process model is available, then process monitoring techniques based on monitoring the model residuals can be employed (Chiang et al., 2001; Davis et al., 2000; Cinar et al., 2007). Simple CPM methods have also been developed that do not require a process model. Control loops that are excessively oscillatory or very sluggish can be identified using correlation or frequency response techniques (Hägglund, 1999; Miao and Seborg, 1999; Tangirala et al., 2005), or by evaluating standard deviations (Rhinehart, 1995; Shunta, 1995). A common problem, control valve stiction, can be detected from routine operating data (Shoukat Choudhury et al., 2008).

Control system performance can be assessed by comparison with a *benchmark*. For example, historical data representing periods of satisfactory control can be used as a benchmark. Alternatively, the benchmark could be an ideal control system performance, such as *minimum variance control*. As the name implies, a minimum variance controller minimizes the variance of the controlled variable when unmeasured, random disturbances occur. This ideal performance limit can be estimated from closed-loop operating data; then the ratio of minimum variance to the actual variance is used as the measure of control system performance. This statistically based approach has been commercialized, and many successful industrial applications have been reported (Kozub, 1997; Desborough and Miller, 2002; Harris and Seppala, 2002; Hoo et al., 2003; Paulonis and Cox, 2003).

Additional information on statistically-based CPM is available in a tutorial (MacGregor, 1988), survey articles (Piovoso and Hoo, 2002; Kourti, 2005), and books (Box and Luceño, 1997; Huang and Shah, 1999; Cinar et al., 2007). Extensions to MIMO control problems, including MPC, have also been reported (Huang et al., 2000; Qin and Yu, 2007; Cinar et al., 2007).

monitoring has also been recognized as a key component of the overall monitoring activity. Process variables are monitored by making simple limit and performance

calculations. Statistical process control (SPC) techniques based on control charts are monitoring techniques widely used for product quality control and other applications where the sampling periods are long relative to process settling times. In particular, Shewhart control charts are used to detect large shifts in mean process behavior, while CUSUM and EWMA control charts are better at detecting small, sustained changes. Multivariate

monitoring techniques such as PCA and PLS can offer significant improvements over these traditional univariate methods when the measured variables are highly correlated. SPC and APC are complementary techniques that can be used together to good advantage. Control performance monitoring techniques have been developed and commercialized, especially methods based on on-line, statistical analysis of operating data.

REFERENCES

- Abnormal Situation Management Consortium (ASM)*, <http://www.asmsortium.com> (2009).
- Alt, F. B., N. D. Smith, and K. Jain, Multivariate Quality Control, in *Handbook of Statistical Methods for Scientists and Engineers*, 2d ed., H. M. Wadsworth (Ed.), McGraw-Hill, New York, 1998, Chapter 21.
- Box, G., and A. Luceño, *Statistical Control by Monitoring and Feedback*, Wiley, New York, 1997.
- Chiang, L. H., E. L. Russell, and R. D. Braatz, *Fault Detection and Diagnosis in Industrial Systems*, Springer, New York, 2001.
- Cinar, A. A. Palazoglu, and F. Kayihan, *Chemical Process Performance Evaluation*, CRC Press, Boca Raton, FL, 2007.
- Davis, J. F., M. J. Piovoso, K. A. Hoo, and B. R. Bakshi, Process Data Analysis and Interpretation, *Advances in Chem. Eng.*, **25**, Academic Press, New York (2000).
- Deming, W. E., *Out of the Crisis*, MIT Center for Advanced Engineering Study, Cambridge, MA, 1986.
- Desborough, L., and R. Miller, Increasing Customer Value of Industrial Control Performance Monitoring—Honeywell's Experience, *Chemical Process Control, CPC-VI*, J. B. Rawlings, B. A. Ogunnaike, and J. Eaton (Eds.), *AIChE Symposium Series*, **98**, 169 (2002).
- Hägglund, T., Automatic Detection of Sluggish Control Loops, *Control Eng. Practice*, **7**, 1505 (1999).
- Harris, T. J., and C. T. Seppala, Recent Developments in Controller Performance Monitoring and Assessment Techniques, *Chemical Process Control, CPC-VI*, J. B. Rawlings, B. A. Ogunnaike, and J. Eaton (Eds.), *AIChE Symposium Series*, **98**, 208 (2002).
- Hoo, K. A., M. J. Piovoso, P. D. Schnelle, and D. A. Rowan, Process and Controller Performance Monitoring: Overview with Industrial Applications, *Int. J. Adaptive Control and Signal Processing*, **17**, 635 (2003).
- Huang, B., R. Kadali, X. Zhao, E. C. Tamayo, and A. Hanafi, An Investigation into the Poor Performance of a Model Predictive Control System on An Industrial CGO Coker, *Control Eng. Practice*, **8**, 619 (2000).
- Huang, B., and S. L. Shah, *Performance Assessment of Control Loops: Theory and Applications*, Springer-Verlag, New York (1999).
- Jackson, J. E. *A User's Guide to Principal Components*, Wiley-Interscience, New York, 1991.
- Jacobs, D. C., Watch Out for Nonnormal Distributions, *Chem. Eng. Progress*, **86**, (11), 19 (1990).
- Jelali, M., An Overview of Performance Assessment Technology and Industrial Applications, *Control Eng. Practice*, **14**, 441 (2006).
- Kourtí, T., Process Analysis and Abnormal Situation Detection: From Theory to Practice, *IEEE Control Systems*, **22** (5), 10 (2002).
- Kourtí, T., Application of Latent Variable Methods to Process Control and Statistical Process Control in Industry, *Int. J. Adaptive Control and Signal Processing*, **19**, 213 (2005).
- Kozub, D. J., Monitoring and diagnosis of chemical processes with automated process control in *Chemical Process Control, CPC-V*, J. C. Kantor, C. E. Garcia, and B. Carnahan, *AIChE Symposium Series*, **93** (No. 316), 83 (1997).
- MacGregor, J. F., On-line Statistical Process Control, *Chem. Eng. Progress*, **84**, (10), 21 (1988).
- MacGregor, J. F., Using On-line Process Data to Improve Quality, *ASQC Statistics Division Newsletter*, **16** (2), 6 (1996).
- Miao, T., and D. E. Seborg, Automatic Detection of Excessively Oscillatory Feedback Control Loops, *Proc. IEEE Internat. Conf. on Control Applications*, 359, Kohala Coast, HI, USA (1999).
- Montgomery, D. C., *Introduction to Statistical Quality Control*, 6th ed., Wiley, New York, 2009.
- Montgomery, D. C., and G. C. Runger, *Applied Statistics and Probability for Engineers*, 4th ed., Wiley, New York, 2007.
- Pande, P. S., R. P. Neuman, and R. R. Cavanagh, *The Six Sigma Way*, McGraw-Hill, New York, 2000.
- Paulonis, M. A., and J. W. Cox, A Practical Approach for Large-Scale Controller Performance Assessment, Diagnosis, and Improvement, *J. Process Control*, **13**, 155 (2003).
- Piovoso, M. J., and K. A. Hoo, Multivariate Statistics for Process Control, *IEEE Control Systems*, **22** (5), 8 (2002).
- Piovoso, M. J., and K. A. (Hoo) Kosanovich, The Use of Multivariate Statistics in Process Control, in *The Control Handbook*, W. S. Levine and R. C. Dorf (Eds.), CRC Press, Boca Raton, FL, 1996, Chapter 33.
- Qin, S. J. and J. Yu, Recent Developments in Multivariable Controller Performance Monitoring, *J. Process Control*, **17**, 221 (2007).
- Rhinehart, R. R., A Watchdog for Controller Performance Monitoring, *Proc. Amer. Control Conf.*, 2239 (1995).
- Romagnoli, J., and M. C. Sanchez, *Data Processing and Reconciliation in Chemical Process Operations*, Academic Press, San Diego, CA, 2000.
- Ryan, T. P., *Statistical Methods for Quality Improvement*, 2d ed., Wiley, New York, 2000.
- Shewhart, W. A., *Economic Control of Quality*, Van Nostrand, New York, 1931.
- Shoukat Choudhury, M. A. A., M. Jain, S. L., and Shah, Stiction—Definition, Modelling, Detection and Quantification, *J. Process Control*, **18**, 232 (2008).
- Shunta, J. P., *Achieving World Class Manufacturing Through Process Control*, Prentice Hall PTR, Englewood Cliffs, NJ, 1995.
- Tangirala, A. K., S. L. Shah, and N. F. Thornhill, PSCMAP: A New Tool for Plant-Wide Oscillation Detection, *J. Process Control*, **15**, 931 (2005).
- Western Electric Company, *Statistical Quality Control Handbook*, Delmar Printing Company, Charlotte, NC, 1956.

EXERCISES

21.1 A standard signal range for electronic instrumentation is 4–20 mA. For purposes of monitoring instruments using limit checks, would it be preferable to have an instrument range of 0–20 mA? Justify your answer.

21.2 An analyzer measures the pH of a process stream every 15 minutes. During normal process operation, the mean and standard deviation for the pH measurement are $\bar{x} = 5.75$ and $s = 0.05$, respectively. When the process is operating normally, what is the probability that a pH measurement will exceed 5.9?

21.3 In a computer control system, the high and low warning limits for a critical temperature measurement are set at the “2-sigma limits,” $\bar{T} \pm 2\hat{\sigma}_T$, where \bar{T} is the nominal temperature and $\hat{\sigma}_T$ is the estimated standard deviation. If the process operation is normal and the temperature is measured every minute, how many “false alarms” (that is, measurements that exceed the warning limits) would you expect to occur during an eight-hour period?

21.4 In order to improve the reliability of a critical control loop, it is proposed that redundant sensors be used. Suppose that three independent sensors are employed and each sensor works properly 95% of the time.

(a) What is the probability that all three sensors are functioning properly?

(b) What is the probability that none of the sensors are functioning properly?

(c) It is proposed that the average of the three measurements be used for feedback control. Briefly critique this strategy.

Hint: See Appendix J for a review of basic probability concepts.

21.5 In a manufacturing process, the impurity level of the product is measured on a daily basis. When the process is operating normally, the impurity level is approximately normally distributed with a mean value of 0.800% and a standard deviation of 0.021%. The laboratory measurements for a period of eight consecutive days are shown below. From an SPC perspective, is there strong evidence to believe that the mean value of the impurity has shifted? Justify your answer.

Day	Impurity (%)	Day	Impurity (%)
1	0.812	5	0.799
2	0.791	6	0.833
3	0.841	7	0.815
4	0.814	8	0.807

21.6 A drought in southern California resulted in water rationing and extensive discussion of alternative water supplies. Some people believed that this drought was the worst one ever experienced in Santa Barbara County.

But was this really true? Rainfall data for a 120-year period are shown in Table E21.6. In order to distinguish between normal and abnormal drought periods, do the following.

(a) Consider the data before the year 1920 to be a set of “normal operating data.” Use these data to develop the target and control limits for a Shewhart chart. Determine if any of the data for subsequent years are outside the chart limits.

(b) Use the data prior to 1940 to construct an s chart that is based on a subgroup of 10 data points for each decade. How many chart violations occur for subsequent decades?

21.7 Develop CUSUM and EWMA charts for the rainfall data of Exercise 21.6 considering the data for 1900 to 1930 to be the “normal operating data.” Use the following design parameters: $K = 0.5$, $H = 5$, $\lambda = 0.25$. Based on these charts, do any of the next three decades appear to be abnormally dry or wet?

21.8 An SPC chart is to be designed for a key process variable, a chemical composition, which is also a controlled variable. Because the measurements are very noisy, they must be filtered before being sent to a PI controller. The question arises whether the variable plotted on the SPC chart should be the filtered value or the raw measurement. Are both alternatives viable? If so, which one do you recommend? (Briefly justify your answers.)

21.9 For the BOD data of Example 21.5, develop CUSUM and EWMA charts. Do these charts indicate an “abnormal situation”? Justify your answer. For the CUSUM chart, use $K = 0.5s$ and $H = 5s$ where s is the sample standard deviation. For the EWMA chart, use $\lambda = 0.25$.

21.10 Calculate the average values of the C_p and C_{pk} capability indices for the BOD data of Example 21.5, assuming that $LSL = 5$ mg/L and $USL = 35$ mg/L. Do these values of the indices indicate that the process performance is satisfactory?

21.11 Repeat Exercise 21.10 for the solids data of Example 21.5, assuming that $USL = 1,600$ mg/L and $LSL = 1,200$ mg/L.

21.12 Consider the wastewater treatment problem of Examples 21.5 and 21.6 and five new pairs of measurements shown below. Calculate the value of Hotelling’s T^2 statistic for each pair using the information for Example 21.6, and plot the data on a T^2 chart. Based on the number of chart violations for the new data, does it appear that the current process behavior is normal or abnormal?

Sample Number	BOD (mg/L)	Solids (mg/L)
1	18.1	1281
2	36.8	1430
3	16.0	1510
4	28.2	1343
5	31.0	1550

Note: The required covariance matrix S in Eq. 21-27 can be calculated using either the *cov* command in MATLAB or the *covar* command in EXCEL.

Table E21.6 Rainfall Data, 1870–1990

Year	Rain (in)	Year	Rain (in)	Year	Rain (in)
1870	10.47	1911	31.94	1951	11.29
1871	8.84	1912	16.35	1952	31.20
1872	14.94	1913	12.78	1953	12.98
1873	10.52	1914	31.57	1954	15.37
1874	14.44	1915	21.46	1955	17.07
1875	18.71	1916	25.88	1956	19.58
1876	23.07	1917	21.84	1957	13.89
1877	4.49	1918	21.66	1958	31.94
1878	28.51	1919	12.16	1959	9.06
1879	13.61	1920	14.68	1960	10.82
1880	25.64	1921	14.31	1961	9.99
1881	15.23	1922	19.25	1962	28.22
1882	14.27	1923	17.24	1963	15.73
1883	13.41	1924	6.36	1964	10.19
1884	34.47	1925	12.26	1965	18.48
1885	13.79	1926	15.83	1966	14.39
1886	24.24	1927	22.73	1967	24.96
1887	12.96	1928	13.48	1968	13.67
1888	21.73	1929	14.54	1969	30.47
1889	21.04	1930	13.91	1970	12.03
1890	32.47	1931	14.99	1971	14.02
1891	17.31	1932	22.13	1972	8.64
1892	10.75	1933	6.64	1973	23.33
1893	27.02	1934	13.43	1974	17.33
1894	7.02	1935	21.12	1975	18.87
1895	16.34	1936	18.21	1976	8.83
1896	13.37	1937	25.51	1977	16.49
1897	18.50	1938	26.10	1978	41.71
1898	4.57	1939	13.35	1979	21.74
1899	12.35	1940	14.94	1980	24.59
1900	12.65	1941	45.71	1981	15.04
1901	15.40	1942	12.87	1982	15.11
1902	14.21	1943	24.37	1983	38.25
1903	20.74	1944	17.95	1984	14.70
1904	11.58	1945	15.23	1985	14.00
1905	29.64	1946	11.33	1986	22.12
1906	22.68	1947	13.35	1987	11.45
1907	27.74	1948	9.34	1988	15.45
1908	19.00	1949	10.43	1989	8.90
1909	35.82	1950	13.15	1990	6.57
1910	19.61				