Part V

STATISTICAL PROCESS MONITORING AND QUALITY CONTROL

Contents

Chapter 1

OVERVIEW AND FUNDAMENTALS OF SPC

List of References

- **Probability Theory**

  1. Papoulis, A. *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, 1991.

- **Statistical Quality / Process Control**

  1. Grant, E. L. & Leavenworth, R. S. *Statistical Quality Control*, 7th Edition, McGraw-Hill, 1996.
  2. Thompson, J. R. & Koronacki, J. *Statistical Process Control for Quality Improvement*, Chapman & Hall, 1993.
  3. Montgomery, D. C. *Introduction to Statistical Quality Control*, Wiley, 1986.

1.1 INTRODUCTION

1.1.1 MOTIVATION FOR SPC

Background:

- Chemical plants are complex arrays of large processing units and instrumentation.
- Even though designed to be steady, operations are in fact very *dynamic*. Perturbations that occur to typical plants can be categorized into two kinds.
  - Normal day-to-day disturbances (e.g., feed fluctuations, heat / mass transfer variations, pump / valve errors, etc.).
  - Abnormal events. (e.g., abnormal feedstock changes, equipment / instrumentation malfunctioning and failures, line leaks and clogs, catalyst poisoning).
- Most of the problems in the former category are handled by automatic control systems.
- Problems in the latter category are relegated to plant operators.

Useful Analogy:

- plant $\rightarrow$ patient
- automatic control systems $\rightarrow$ patient's immune system
- operator $\rightarrow$ doctor
- DCS, plant information data base $\rightarrow$ modern medical diagnosing instruments.

Key Point: Why SPC?

- Measurements often contain statistical errors / outliers + There are inherent randomness in the process. ⇒ Not easy for operators to distinguish between the normal and abnormal conditions (until the problems fully develop to produce undesirable consequences)
- SPC in the tranditional sense of the word is a diagnostic tool, an indicator of quality problems. However, it does not identify the source of the problem, nor corrective action to be taken.

Comment:
Advances in sensors / information technology made it possible to access an enormous amount of information coming from the plant. The bottleneck these days is not the amount of information, but the operator's ability to process them. We need specialists (e.g., statisticians or SPC techniques) who can pre-process these information into a form useful to the operators.

1.1.2 MAIN POINTS

Traditional SPC techniques have limited values in process industries because

- they assume time-independence of measurements during normal (*in-control*) periods of operation.
- they assume that control should be done only when unusual events occur, i.e.,
  - the costs of making adjustements are very high.
  - little incentive for quality control during normal (*in-control*) situations.

These assumptions are usually not met in the process industries.

In this lecture, we will extend the tradtional methods to remove these deficiencies. The end result will be an *integrated approach to statistical monitoring, prediction and control.*

1.2 TRADITIONAL SPC TECHINQUES

Most traditional techniques are *static*, *univariate* and *chart-based.* We briefly review some of the popular techniques.

1.2.1 Milestones / Key Players of SQC

- **Pareto's Maxim [1848-1923]:**

  > Catastrophically many failures in a system are often attributable to only a small numer of causes. General malaise is seldom the root problem.

  In order to improve the system, a skiled investigator must find Pareto's glitches and correct them.

- **Shewart's Hypothesis [1931]:** Quantitized the Pareto's qualitative observations using a mixture of distributions. Developed control charts to identify *out-of-control* epochs, which enables one to discover the systemic cause of *Pareto's glitches* by backtracking. The result is a kind of step-wise optimization of a system.

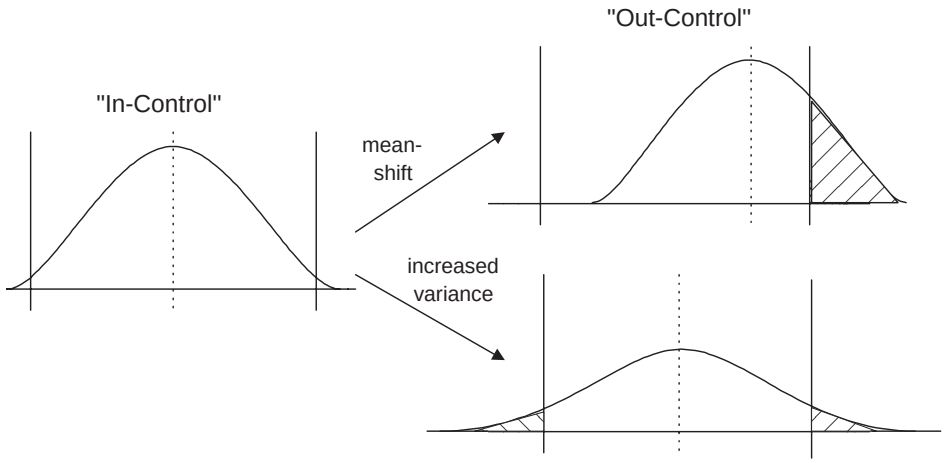
- **Deming's Implementation [1950s-1980s]:** Deming implemented the SPC concept in Japan after Second World War II (and later in North America) on a massive scale. The result was that Japan moved

quickly into a position to challenge and then surpass American automobile and electronic equipments production.

1.2.2 SHEWART CHART

Basic Idea:

The basic idea is that there are switches in time which transfer the generating process into a distribution not typical of the dominant distribution. These switches manifest themselves into different average measurements and variances of the products.



Procedure

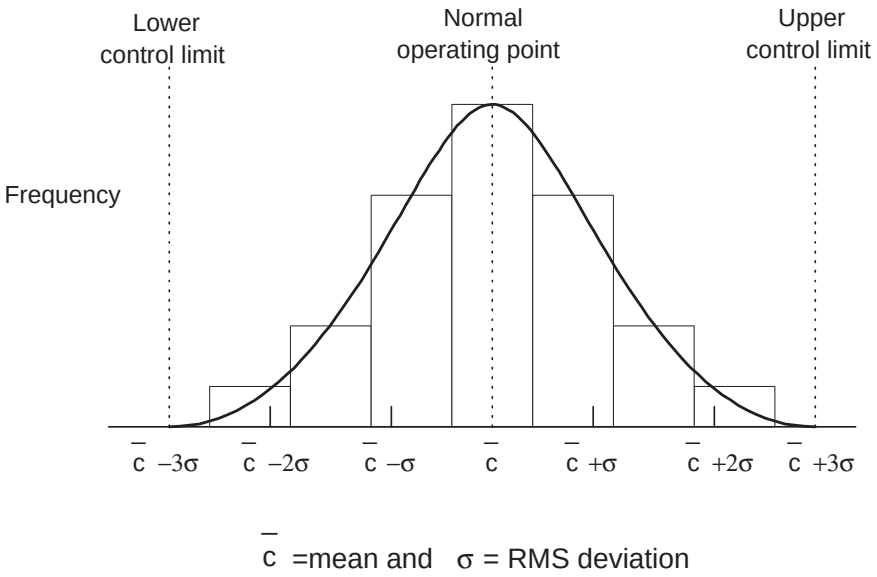
The procedure for constructing and using the Shewart chart is as follows:

- Collect samples during normal (*in-control*) epochs of the operation.
- Compute the sample mean and standard deviation.

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N}\hat{y}(i) \tag{1.1}$$

$$\sigma_y = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}(i)-\bar{y})^2} \tag{1.2}$$

You may also want to construct the histogram to see if the distribution looks close to being normal.

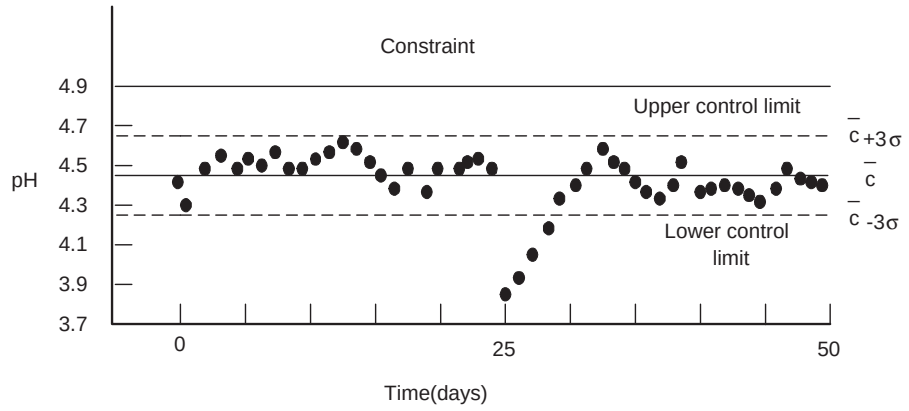


- Establish control limits based on the desired probability level γ. For normal distribution,

b	$P_r\{(\frac{y-\bar{y}}{\sigma_y})^2 \leq b\}$
0.02	0.1
1	0.632
2.71	0.9
3.84	0.95
5.02	0.975
6.63	0.99
7.88	0.995
9.0	0.997

For instance, the bound $\bar{y} \pm 3\sigma_y$ corresponds to the 99.7% probability level.

- Plot the quality data against the established limits. If data violate the limit (repeatedly), *out of control* is declared.



Assessment

The typicallly used 3 σ bound is thought to be both too small and too large, i.e.,

- too small for rejecting outliers (i.e., to prevent false alarms).
- too large for (quickly) catching small mean-shifts and other time-correlated changes.

The first problem can be solved by using the following modification.

Modification: "q-In-A-Row" Implementation

A useful modification is the *q-in-a-row* implementation. In this case,

> *out-of-control* is declared only when the bound is violated q times in a row.

q is typically chosen as 2-3.

Note that, assuming *samples during normal operation are independent in time*, the probability of samples being outside the γ probability bound q consecutive times during the *in-control* period is $(1-\gamma)^q$. For instance, with $\gamma = .9$ and $q = 3$, $(1-\gamma)^q = 0.001$. Hence, in effect, one is using 99.9% confidence limit.

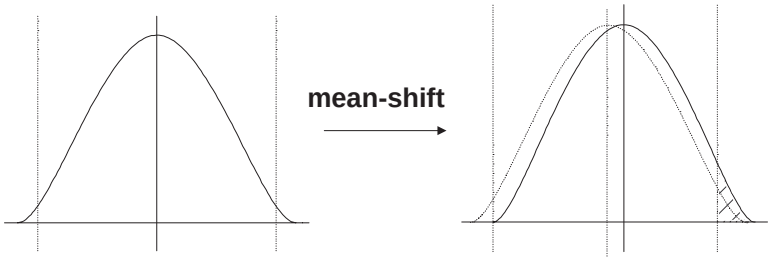
This way, the bound can be tightened without dropping the probability level (or the probability level can be raised without enlarging the bound). The q-in-a-row concept is very effective in rejecting measurement outliers or other short-lived changes for an obvious reason.

On the other hand, with a large q, the detection time is slowed down. In addition, the above concept relies on the fact that *sample data during normal operation are independent.* This may not be true. In this case, false alarms can result using a bound computed under the assumption of independence.

1.2.3 CUSUM CHART

Basic Idea

As shown earlier, with the Shewart Chart, one may have difficulty (quickly) detecting *small, but signficant* mean shifts.



To alleviate this problem, the Shewart Chart can be used in conjunction

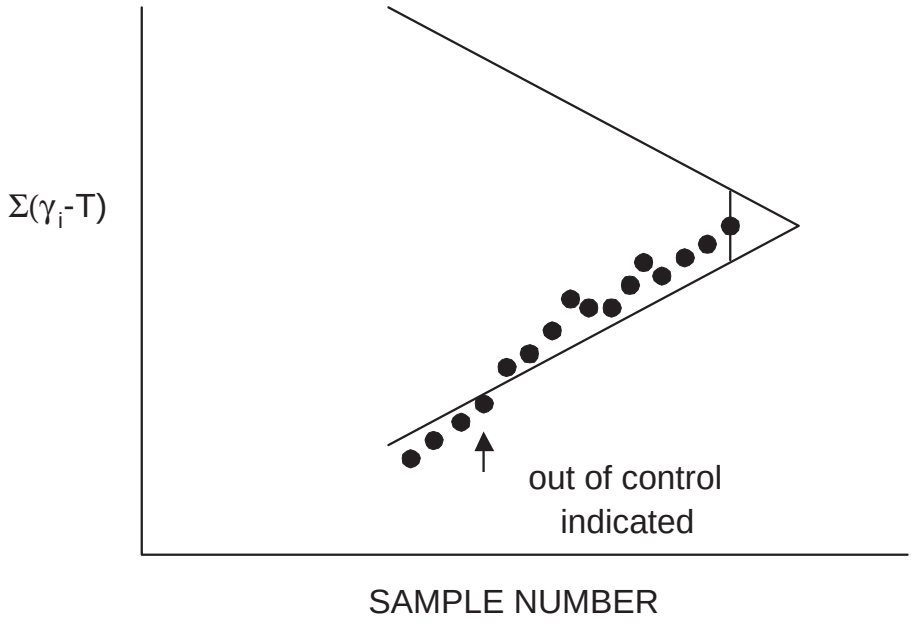
with the CUSUM Chart. The CUSUM chart plots the following:

$$s(k) = \sum_{i=1}^{k} (\hat{y}(i) - \bar{y}) = s(k-1) + (\hat{y}(i) - \bar{y}) \tag{1.3}$$

DuPont has more than 10,000 CUSUM charts currently being used.

Graphical Procedure

In this procedure, $s(k)$ is plotted on a chart and monitored. Any mean shift should show up as a change in the slope in the CUSUM plot. To test the statistical signficance of any change, a V mask is often employed:



Here one can set some tolerance on the rate of rise / fall. If any leg of the V-mask crosses the plotted CUSUM values, a *mean-shift* is thought to have occurred.

Non-Graphical Procedure

In practice, rather than using the graphical chart, one computes the two

cumulative sums

$$s(k) = \max\{0, s(k-1) + (y(k) - \bar{y}) - \epsilon)\} \quad (1.4)$$

$$t(k) = \min\{0, t(k-1) + (y(k) - \bar{y}) + \epsilon)\} \quad (1.5)$$

where ϵ is the allowable slack.

One then tests:

$$\text{whether } s(k) \geq h \text{ or } h(k) \leq -h$$

If either is true, *out-of-control* is declared.

Two parameters need to be decided:

- The allowed slack ϵ is chosen as one-half of the smallest shift in the mean onsidered to be important.
- h is chosen as the compromise that will result in an acceptable long average run length (ARL) in normal situations, but an acceptably short ARL when the process mean shifts as much as 2ϵ units.

1.2.4 EWMA CHART

Basic Idea

The exponentially weighted moving average control chart plots the following exponentially weighted moving average of the data:

$$\begin{aligned} z(i) &= (1-\alpha)(y(i) + \alpha y(i-1) + \alpha^2 y(i-2) + \cdots\cdots + \alpha^k y(0)) \\ &= (1-\alpha)y(i) + \alpha z(i-1) \end{aligned} \quad (1.6)$$

Procedure

- As before, collect operational data from the *in-control* epochs.

- Establish its mean and standard deviation. This gives the target line and upper / lower bounds for the EWMA Chart.
- Compute on-line the sequence $z(k)$ according to $z(k) = y(k) + \alpha z(k-1)$ and plot it on the chart. As before, *out of control* is declared when either of the bounds is exceeded.

Advantages / Justification

The advantages of the EWMA are as follows:

- Note that the above amounts to first-order filtering of the data. The sensitivity to measurement outliers and other high-frequency (short-lived) variations is thus reduced.
- With $\alpha = 0$, one gets the Shewart Chart. As $\alpha \to 1$, it approaches the CUSUM chart. Hence, the parameter α affords the user some flexibility. (The choice of $\alpha = 0.8$ is the most common in practice.)
- This filtering is thought to provide one-step ahead prediction of $y(k)$ in many situations (more on this later).

On the other hand, one does lose some high frequency information through filtering, so it can slow down the detection.

1.3 MULTIVARIATE ANALYSIS

1.3.1 MOTIVATION

Main Idea / Motivation

A good way way to speed up the detection of abnormality and reduce the frequency of false alarm is **utilize more measurements**. This may mean

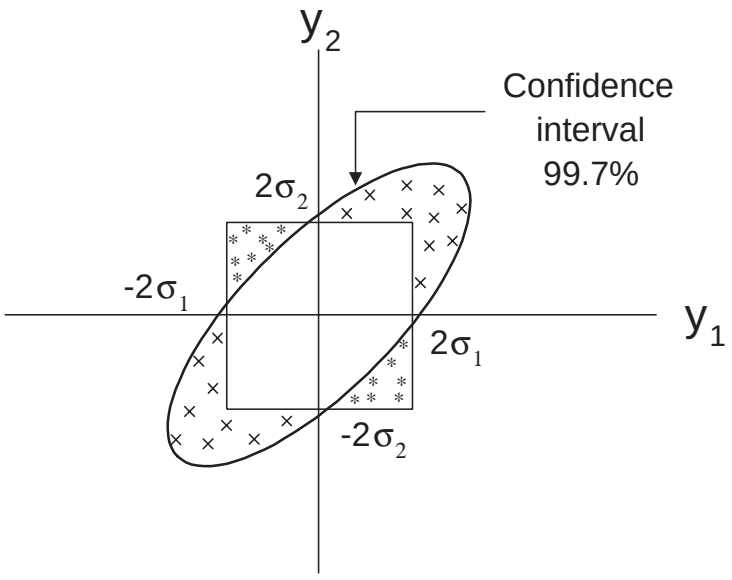
- simultaneous analysis of several quality variables
- including other (more easily and efficiently measured) process variables into the monitoring.

Very Important Point

These measurements are often not independent, but carry significant correlation. It is important to account for the existing correlation through the use of multivariate statistics.

Motivating Example

Let us demonstrate the importance of considering the correlation through the following simple example. Assume that we are monitoring two outputs y_1 and y_2 and their underlying probability distribution is jointly normal. If the correlation is strong, the data distribution and joint-confidence interval looks as below:



Considering the two measurements to be independent results in the conclusion that the probability of being outside the box is approximately $(1 - 0.95)^2 \approx 0.03$. The problems are:

- There are points (marked with $*$ in the above) that are outside the probability level γ, but fall well within the two σs on both univariate charts. This means missed faults.
- There are points (marked with $\times$) that are inside the joint confidence interval of 99.7% probability level, but are outside the box. This means false alarms.

Conclusions:

- The most effective thing to do is to establish an elliptical confidence interval corresponding to a desired probability level γ and see if the measurement falls outside the interval.
- *q-in-a-row* concept can be utilized as before, if desired.
- On the other hand, as the dimension of the output rises, graphical inspection is clearly out of question. It is desirable to reduce the variables into one variable that can be used for a monitoring purpose.

1.3.2 BASICS OF MULTIVARIABLE STATISTICS AND CHI-SQUARE MONITORING

Computation of Sample Mean and Covariance

Let y be a vector containing n variables:

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \tag{1.7}$$

Then the sample mean and covariance can be defined as before:

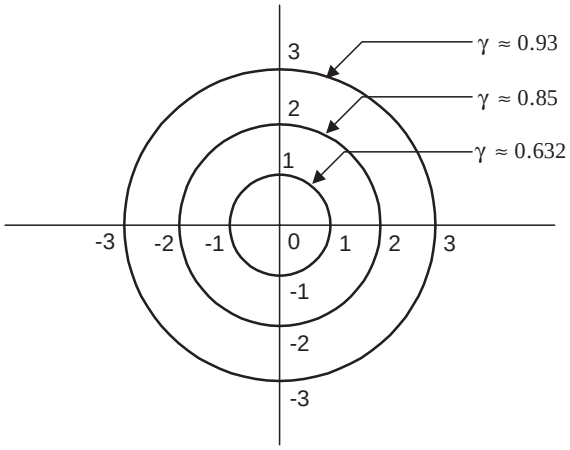
$$\bar{y} = \begin{bmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_n \end{bmatrix} = \frac{1}{N} \sum_{i=1}^{N} \begin{bmatrix} y_1(i) \\ \vdots \\ y_n(i) \end{bmatrix} \tag{1.8}$$

$$R_y = \frac{1}{N} \sum_{i=1}^{N} \left\{ \left(\begin{bmatrix} y_1(i) \\ \vdots \\ y_n(i) \end{bmatrix} - \begin{bmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_n \end{bmatrix} \right) \left(\begin{bmatrix} y_1(i) \\ \vdots \\ y_n(i) \end{bmatrix} - \begin{bmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_n \end{bmatrix} \right)^T \right\} \tag{1.9}$$

As $N \to \infty$, the above should approach the mean and covariance (assuming stationarity). Hence, N should be fairly large for the above to be meaningful.

Decorrelation & Normalization: For Normally Distributed Variables

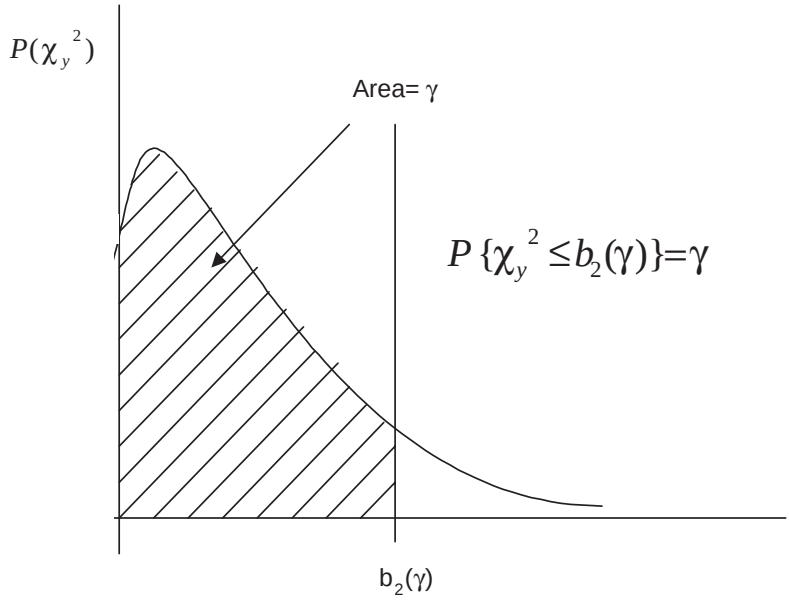
Assuming the underlying distribution is *normal*, the distribution of $z \triangleq R_y^{-1/2}(y - \bar{y})$ is normal with zero mean and identity covariance matrix. Hence, $R_y^{-1/2}$ can be interpreted as a transformation performing both decorrelation and normalization.The distribution for the two-dimensional case looks as below:



Chi-Square Distribution

Hence, the following quantity takes on the chi-square distribution of degree-of-freedom n:

$$\chi_y^2 \triangleq z^T z = (y - \bar{y})^T R_y^{-1} (y - \bar{y}) \tag{1.10}$$



Distribution of $\chi_y^2 = z^T z = z_1^2 + z_2^2$

For any given probability level γ, one can establish the elliptical confidence interval

$$(y - \bar{y})^T R_y^{-1} (y - \bar{y}) \leq b_n(\gamma) \tag{1.11}$$

simply by reading off the values $b_n(\gamma)$ from a chi-square value table.

Chi-square percen $\chi_u^2(n)$

n \ u	0.005	0.01	0.025	0.05	0.1	0.9	0.95	0.975	0.99	0.995
1	0	0	0	0	0.02	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.1	0.21	4.61	5.99	7.38	9.21	10.6
3	0.07	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.34	12.84
4	0.21	0.3	0.48	0.71	1.06	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.2	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.7	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.6	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	4.4	5.23	6.3	18.55	21.03	23.34	26.22	28.3
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.6	5.23	6.26	7.26	8.55	22.31	25	27.49	30.58	32.8
16	5.14	5.81	6.91	7.69	9.31	23.54	26.3	28.85	32	34.27
17	5.7	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	27.2	30.14	32.85	36.91	38.58
20	7.42	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40
22	8.6	9.5	11	12.3	14	30.8	33.9	36.8	40.3	42.8
24	9.9	10.9	12.4	13.8	15.7	33.2	36.4	39.4	43	45.6
26	11.2	12.2	13.8	15.4	17.3	35.6	38.9	41.9	45.6	48.3
28	12.5	13.6	15.3	16.9	18.9	37.9	41.3	44.5	48.3	51
30	13.8	15	16.8	18.5	20.6	40.3	43.8	47	50.9	53.7
40	20.7	22.2	24.4	26.5	29.1	51.8	55.8	59.3	63.7	66.8
50	28	29.7	32.4	34.8	37.7	63.2	67.5	71.4	76.2	79.5

For $n > 50$: $\chi_u^2(n) = \frac{1}{2}(z_u + \sqrt{2n-1})^2$

Now one can simply monitor $\chi_y^2(k)$ against the established bound.

Limitations of Chi-Square Test

The chi-square monitoring method that we discussed has two drawbacks.

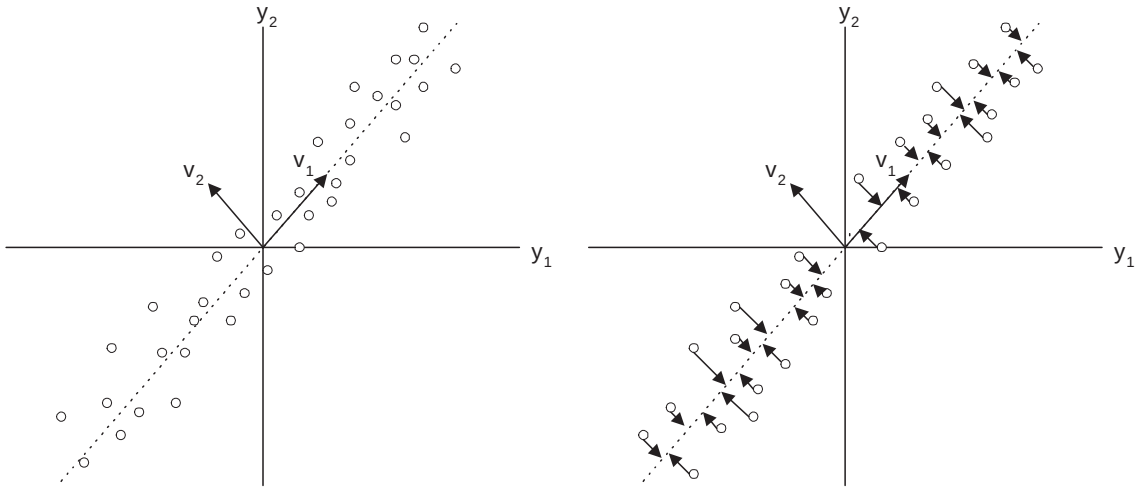
- *No Useful Insight for Diagnosis*
  Although the test suggests that there may be an abnormality in the operation, it does not provide any more insight. One can store all the output variables and analyze their behavior whenever an abnormality is indicated by the chi-square test. However, this requires a large storage space and analysis based on a large correlated data set is anything but a difficult, cumbersome task.

- *Sensitivity to Outliers and Noise*
  Note that the variables are normalized through $R_y^{-1/2}$. For an ill-conditioned R_y, gains of very different magnitudes are applied to different combinations of the y elements in the normalization process. This can cause extreme sensitivity to noise, outliers, etc.

1.3.3 PRINCIPAL COMPONENT ANALYSIS

A solution to the both problem is to monitor and store only the *principal components* of the output vector.

What's The Idea?
Consider the following two-dimensional case:



It is clear that, through an appropriate coordinate transformation, one can explain most of the variation with a single variable.

The SVD of the covariance matrix provides a useful insight for doing this. For the above case, the SVD looks like

$$R_y = \begin{bmatrix} v_1 & v_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^t \end{bmatrix}, \qquad \sigma_1 \gg \sigma_2$$

Computing the Princial Components: Using SVD

The principal components may be computed using the singular value decomposition of R_y as follows:

$$R_y = \left[\begin{array}{ccc|ccc} v_1 & \cdots & v_m & v_{m+1} & \cdots & v_n \end{array}\right] \left[\begin{array}{ccc|ccc} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_m & & & \\ \hline & & & \sigma_{m+1} & & \\ & & & & \ddots & \\ & & & & & \sigma_n \end{array}\right] \left[\begin{array}{c} v_1^T \\ \vdots \\ v_m^T \\ \hline v_{m+1}^T \\ \vdots \\ v_n^T \end{array}\right] \tag{1.12}$$

One can, for instance, choose m such that

$$\frac{\sum_{i=1}^{m} \sigma_i}{\sum_{i=1}^{n} \sigma_i} \geq \gamma \tag{1.13}$$

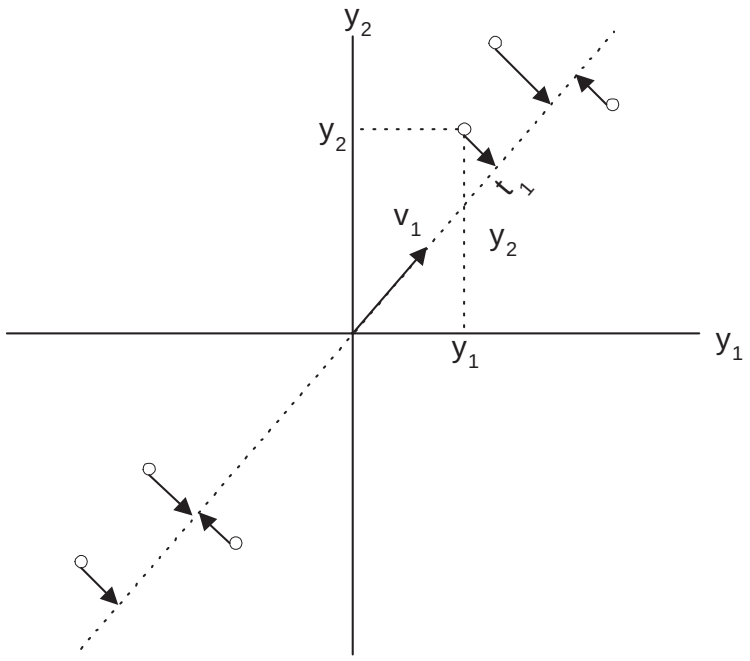
where γ is the tolerance parameter close to 1 (say .99), or such that

$$\sigma_m \gg \sigma_{m+1} \tag{1.14}$$

Usually, $m \ll n$.

$v_1, \cdots, v_m$ are called *principal component directions*. Define the *score variables* for the principal component directions as

$$t_i = v_i^T y, \qquad i = 1, \cdots, m \tag{1.15}$$



These score variables are independent of one another since

$$
\begin{aligned}
E\left\{ \begin{bmatrix} t_1 \\ \vdots \\ t_m \end{bmatrix} \begin{bmatrix} t_1 \\ \vdots \\ t_m \end{bmatrix}^T \right\} &= E\left\{ \begin{bmatrix} v_1^T \\ \vdots \\ v_m^T \end{bmatrix} (y(k)-\bar{y})(y(k)-\bar{y})^T \begin{bmatrix} v_1 & \cdots & v_m \end{bmatrix} \right\} \\
&= \begin{bmatrix} v_1^T \\ \vdots \\ v_m^T \end{bmatrix} R_y \begin{bmatrix} v_1 & \cdots & v_m \end{bmatrix} \\
&= \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_m^2 \end{bmatrix} \qquad (1.16)
\end{aligned}
$$

Example

Show a 4-dimensinal case, perform SVD and explain what it means.
Actually generate a large set of data and show projection to each principal component direction.

- Generate 1000 data points (from the normal distribution).
- Plot each y (time vs. value plot for each variable).
- Compute the sample mean and covariance.
- Perform SVD of the sample covariance.
- Compute principal components.
- Plot each variable along with its prediction from the two principal components ($\hat{y} = t_1 v_1 + t_2 v_2$)

Monitoring Based on Principal Component Analysis

- t_i's are perfect candidates for monitoring since they are: (1) independent of one another, and (2) relatively low in dimension.
- The residual vector can be computed as

$$r(k) = y(k) - \sum_{i=1}^{m} \underbrace{(v_i^T y(k))}_{t_i(k)} v_i \tag{1.17}$$

  The above residual vector represents the contribution of the parts that were thrown out because their variations were judged to be insignificant from the normal operating data. The size of the residual vector should be monitored in addition, since a significant growth in its size can indicate an abnormal (*out-of-control*) situation.

Advantages

The advantage of the two-tier approach is that one can gain much more information from the monitoring. Often times, when the monitoring test indicates a problem, useful additional insights can be gained by examining

- the direction of the principal component(s) which has violated the bound
- the residual vector if its size has gone over the tolerance level.

1.3.4 EXAMPLE: MULTIVARIATE ANALYSIS VS. SINGLE VARIATE ANALYSIS

Compare Univariate vs. Multivariate. Compare chi-square test vs. PC monitoring.

1.4 TIME SERIES MODELING

1.4.1 LIMITATIONS OF THE TRADITIONAL SPC METHODS

In the process industries, there are two major sources for quality variances:

- Equipment / instrumentation malfunctioning.
- feed variations and other disturbances.

Usually, for the latter, the dividing line between normal and abnormal are not as clear-cut since

- they occur very often.
- they tend to fluctuate quite a bit from one time to another (but with strong temporal correlations).
- they often cannot be eliminated at source.

Because of the frequency and nature of these disturbances, they cannot be classified as *Pareto's glitches* and *normal* periods (*in-control* epochs) must be defined to include their effects.

The implications are

- Quality variances during what are considered to be *normal* periods of operation can be quite large.
- Process and quality variables show significant *time* correlations (due to the process dynamics as well as the temporal correlations in these disturbances).
- There is significant incentive to reduce the variance even during *in-control* periods through adjustment of process condition (*e.g.*, temperature, pressure).

These considerations point to the following potential shortcomings of the traditional approaches:

- *Testing For a Wrong Hypothesis?*
  The previously-discussed traditional methods can be considered as a kind of hypothesis testing. The hypothesis tested is:

  > During *in-control* epochs (i.e., periods of normal operations), the measured variable is serially independent with the mean and variance corresponding to the chosen target and the bound.

  While the above hypothesis is reasonable in many industries like the automotive industries and parts industries where SQC has proven invaluable, *its validity is highly questionable for chemical and other process industries.* As mentioned above, in these industries, measured variables exhibit significant time correlation even in normal (*in-control*) situations.
- *Lack of Control Need During In-Control Period?*
  SPC is based on the assumption that control adjustments should be made only when an abnormal (*out-of-control*) situation arises. This is

sensible for many industries (e.g, automotive industries) where there are assignable causes to be removed. In reality, quality variables in many chemical processes show significant variances even during *in-control* periods. The deviations are usually time-correlated giving opportunities for control through process input adjustments (which can lower the quality variance, leading to economic savings, more consistent products and less environmental problems, etc.). In most cases, little costs are associated with adjustments.

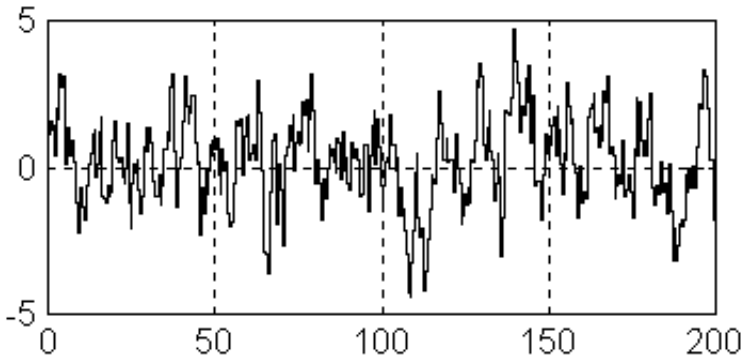
For the remainder, we will highlight the above limitations through simple examples and propose remedies / alternatives.

1.4.2 MOTIVATING EXAMPLE

The Simple First-Order Case.
To understand the limitation arising from ignoring the time correlation, consider the situation where the output follows the pattern

$$(y(k) - \bar{y}) = \alpha(y(k-1) - \bar{y}) + \varepsilon(k) \tag{1.18}$$

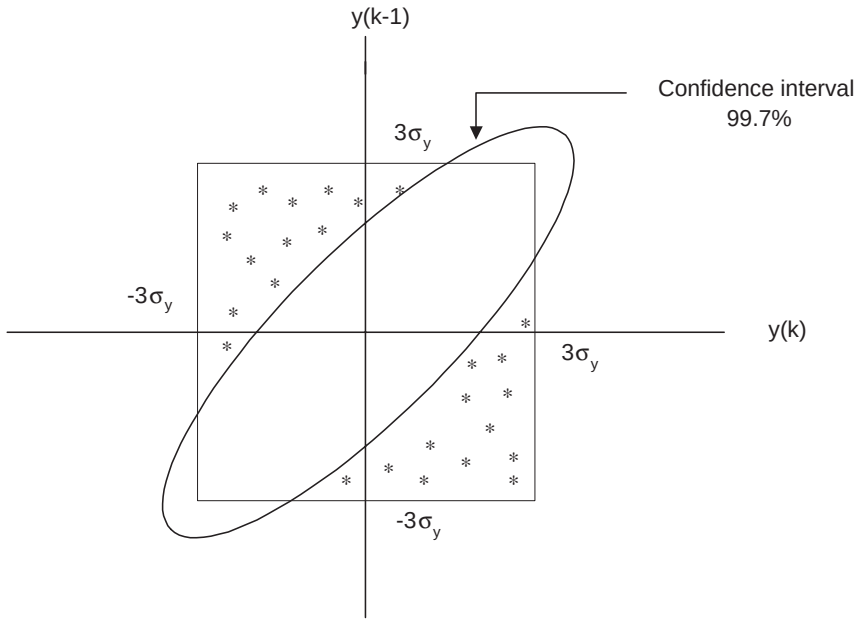
where $\varepsilon(k)$ is an independent (white) sequence with zero mean and variance σ_ε^2. A plot of an example sequence is shown below.



Note that

$$
\begin{aligned}
& E\left\{\begin{bmatrix} y'(k) \\ y'(k-1) \end{bmatrix}\begin{bmatrix} y'(k) & y'(k-1) \end{bmatrix}\right\} \\
= \; & E\left\{\begin{bmatrix} \alpha y'(k-1)+\varepsilon(k) \\ y'(k-1) \end{bmatrix}\begin{bmatrix} \alpha y'(k-1)+\varepsilon(k) & y'(k-1) \end{bmatrix}\right\} \\
= \; & \begin{bmatrix} \alpha^2\sigma_y^2+\sigma_\varepsilon^2 & \alpha\sigma_y^2 \\ \alpha\sigma_y^2 & \sigma_y^2 \end{bmatrix} \qquad (1.19)
\end{aligned}
$$

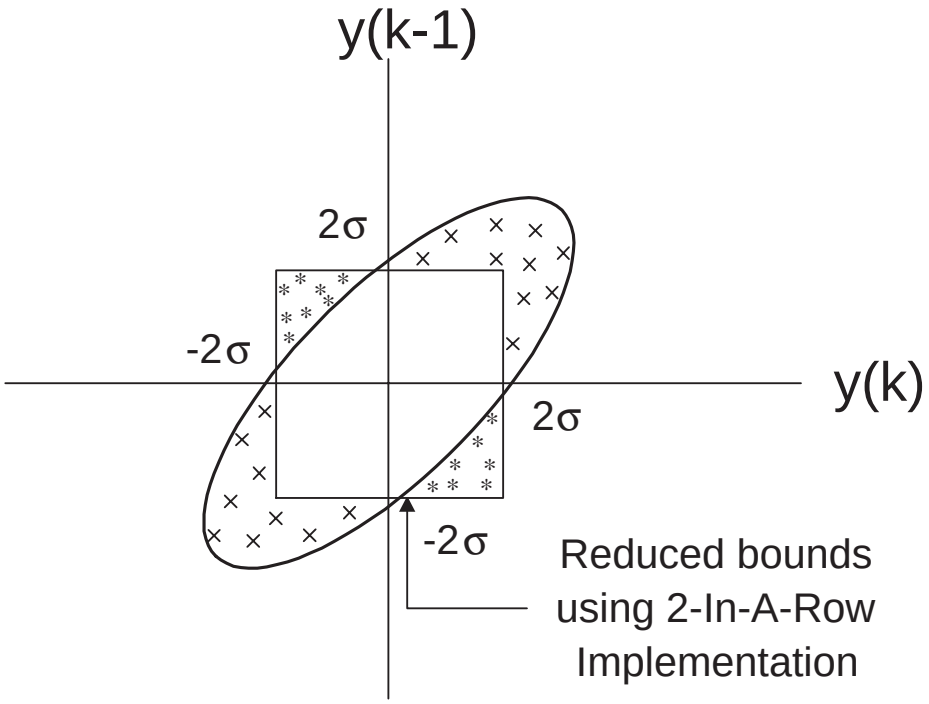
The plot of a confidence interval may look as below:



What's The Problem?

Note that, if $y'(k)$ is monitored through the Shewart chart, one would not be able to catch points marked with * that are outside the 99.7% confidence interval (missed faults).

In order to catch these points, one may choose to tighten the bounds. However, doing this may cause false alarms.



In short, by ignoring the time correlation, one gets bounds that are inefficient.

A Solution?

One solution is to model the sequence using a time series and decorrelate the sequence. For instance, one may fit to the data a time-series model of form

$$y'(k) = a_1 y'(k-1) + \varepsilon(k) \tag{1.20}$$

where $\varepsilon(k)$ is a white (time-independent) sequence. The one-step-ahead prediction based on the above model is

$$\hat{y}'(k|k-1) = a_1 y'(k-1) \tag{1.21}$$

The prediction error is

$$y'(k) - \hat{y}'(k|k-1) = (\alpha - a_1) y'(k-1) + \varepsilon(k) \tag{1.22}$$

Note that, assuming the model parameter matched the true value ($a = \alpha$), the prediction error is $\varepsilon(k)$ which is an independent sequence. Hence, the idea goes as follows:

> Apply the statistical monitoring methods to the prediction error sequence $\varepsilon(k)$, since it satisfies the basic assumption of independence underlying these methods.

1.4.3 TIME-SERIES MODELS

Assuming the underlying distribution is stationary (during *in-control* periods), one can use a time series to model the temporal correlation.

Various Model Types

Different forms of time series models exist for modeling correlation of a time sequence. We will drop the notation $'$ and use y to represent y' for simplicity. The following is an Auto-Regressive (AR) model of order n:

$$y(k) = a_1 y(k-1) + a_2 y(k-2) + \cdots + a_n y(k-n) + \varepsilon(k) \tag{1.23}$$

The parameters can be obtained using the linear least squares method. A more complex model form is the following Auto-Regressive Moving Average (ARMA) model:

$$\begin{aligned} y(k) \;=\; & a_1 y(k-1) + a_2 y(k-2) + \cdots + a_n y(k-n) \\ & + \varepsilon(k) + c_1 \varepsilon(k-1) + c_n \varepsilon(k-n) \end{aligned} \tag{1.24}$$

The above model structure is more general than the AR model, and hence much fewer terms (i.e., lower n) can be used to represent the same random sequence. However, the parameter estimation problem this time is nonlinear. Often, pseudo-linear regression is used for it.

General Model Form

The general form of a model for a stationary sequence is

$$y(k) = H(q^{-1})\varepsilon(k) \tag{1.25}$$

where $H(q^{-1})$ can be interpreted as a filter with $H(0) = 1$. For instance, for AR model,

$$H(q^{-1}) = \frac{1}{1 - a_1 q^{-1} - \cdots - a_n q^{-n}} \tag{1.26}$$

For ARMA model,

$$H(q^{-1}) = \frac{1 + c_1 q^{-1} + \cdots + c_n q^{-n}}{1 - a_1 q^{-1} - \cdots - a_n q^{-n}} \tag{1.27}$$

1.4.4 COMPUTATION OF PREDICTION ERROR

Key Idea

The one-step-ahead prediction based on model (1.25) is

$$\hat{y}(k|k-1) = E\{y(k)|y(k-1), \cdots\} = (1 - H^{-1}(q^{-1}))y(k) \tag{1.28}$$

Note that, since the constant term cancels out in $(1 - H^{-1}(q^{-1}))$ (because $H(0) = 1$), it contains at least one delay and the RHS does not requires $y(k)$.

The optimal prediction error is

$$y(k) - \hat{y}(k|k-1) = H^{-1}(q^{-1})y(k) = \varepsilon(k) \tag{1.29}$$

Hence, assuming the model correctly represents the underlying system, the prediction error should be white. In conclusion,

> Compute $\varepsilon(k) = H(q^{-1})y(k)$ and monitor $\varepsilon(k)$ using the Shewart's method, etc.

by filtering the output sequence with filter $H^{-1}(q^{-1})$, one can decorrelate

the sequence, which results in an independent variable suitable for traditional SPC methods. $H^{-1}(q^{-1})$ can be thought of as *whitening filter*.

1.4.5 INCLUDING THE DETERMINSTIC INPUTS INTO THE MODEL

What's the Issue?

In many cases, variation of the output variable may not be entirely stochastic. That is, there may be deterministic inputs that contribute to the observed output behavior. Included in the deterministic inputs are

- measured disturbances
- manipulated inputs (e.g., setpoints to the existing loops)

In this case,

> rather than viewing the output behavior as being entirely stochastic, we can add the effect of the determinstic input explicitly into the model for a better prediction.

Another option is to include them in the output vector. However, the former option is more convenient for designing a supervisory control system.

Model Form : Same As Before But With Extra Inputs

In the linear systems context, one may use a model of the form

$$y(k) = G(q^{-1})u(k) + H(q^{-1})\varepsilon(k) \tag{1.30}$$

For instance, the following is the ARX (Auto-Regressive with eXtra input) model.

$$\begin{aligned} y(k) \;=\; & a_1 y(k-1) + a_2 y(k-2) + \cdots + a_n y(k-n) \\ & + b_1 u(k-1) + \cdots + b_m u(k-m) + \varepsilon(k) \end{aligned} \tag{1.31}$$

Another popular choice is the ARMAX (Auto-Regressive Moving Average with eXtra input) model, which looks like

$$\begin{aligned} y(k) &= a_1 y(k-1) + a_2 y(k-2) + \cdots + a_n y(k-n) \\ &\quad + b_1 u(k-1) + \cdots + b_m u(k-m) \\ &\quad + \varepsilon(k) + c_1 \varepsilon(k-1) + \cdots + c_n \varepsilon(k-n) \end{aligned} \tag{1.32}$$

Monitoring: No More Difficult!

The one-step-ahed prediction is given as

$$y(k|k-1) = G(q^{-1})u(k) + (I - H^{-1}(q^{-1}))(y(k) - G(q^{-1})u(k)) \tag{1.33}$$

and the prediction is once again

$$y(k) - y(k|k-1) = H^{-1}(q^{-1})(y(k) - G(q^{-1})u(k)) = \varepsilon(k) \tag{1.34}$$

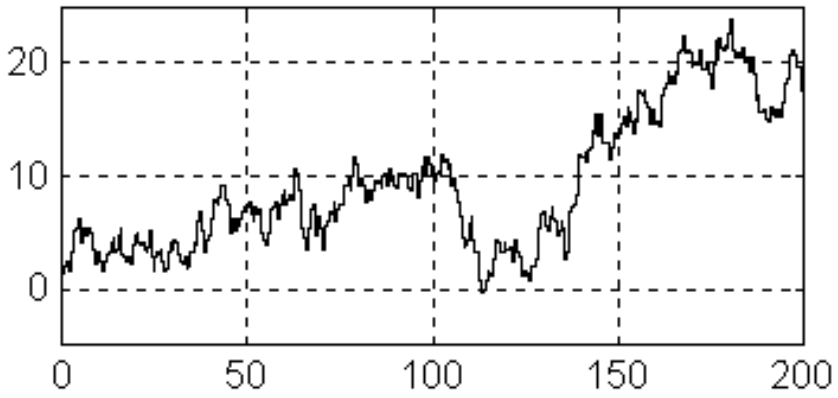
Control Opportunity: Additional Benefit

Having the determinstic inputs in the model also give opportunities to control the process (in addition to the monitoring). That is, one can manipulate the deterministic input sequence $u(k)$ to shape the output behavior in a desirable manner (e.g., no bias, minimum-variance).

1.4.6 MODELING DRIFTING BEHAVIOR USING A NONSTATIONARY SEQUENCE

Basic Problem

For many processes, even during what is considered to be *in-control* periods, variables do not have a fixed mean or level, but rather are drifting / *mean-shifting* (*nonstationary*). Shown below is an example of such a process:



Model Form

the following ARIMA model is the popular choice:

$$\begin{aligned}(1 - a_1 q^{-1} - \cdots - a_n q^{-n}) y(k) &= (b_1 q^{-1} + \cdots + b_m q^{-m}) u(k) \\ &+ (1 + c_1 q^{-1} + \cdots + c_n q^{-n}) \underbrace{\frac{1}{1 - q^{-1}}}_{\text{integrator}} \varepsilon(k)\end{aligned} \tag{1.35}$$

The above can be re-expressed as:

$$\begin{aligned}\Delta y(k) &= a_1 \Delta y(k-1) + a_2 \Delta y(k-2) + \cdots + a_n \Delta y(k-n) \\ &+ b_1 \Delta u(k-1) + \cdots + b_m \Delta u(k-m) \\ &+ \varepsilon(k) + c_1 \varepsilon(k-1) + c_n \varepsilon(k-n)\end{aligned} \tag{1.36}$$

Hence, simple differencing of input and output data gets rid of the stationarity.

More generally, a model for a nonstationary sequence takes the form of

$$\Delta y(k) = G(q^{-1}) \Delta u(k) + H(q^{-1}) \varepsilon(k) \tag{1.37}$$

Decorrelation: Just Needs Differencing!

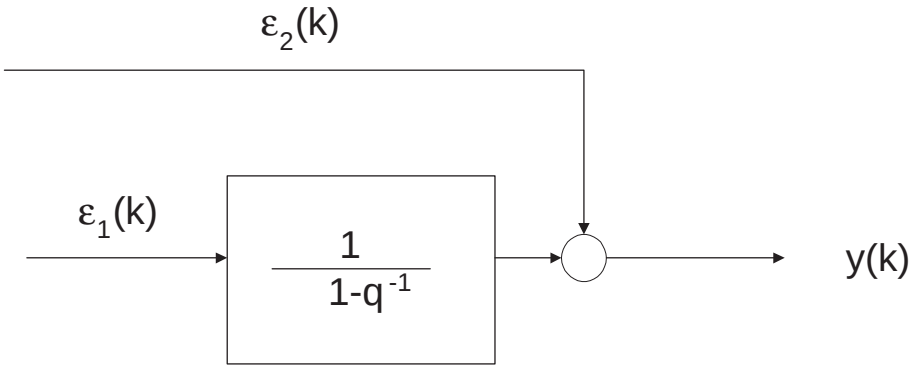
Once again, de-correlation can be done through

$$\varepsilon(k) = H^{-1}(q^{-1})(\Delta y(k) - G(q^{-1})\Delta u(k)) \tag{1.38}$$

Hence, the only extra thing required is differencing of the data, both in prior to the model construction and decorrelation through filtering with the model inverse.

Example

Consider the case where the output is sum of the following two random effects:



Then, the overall behavior of y can be expressed as

$$y(k) = \frac{1 - \alpha q^{-1}}{1 - q^{-1}}\varepsilon(k) \tag{1.39}$$

There also is a result that all nonstationary disturbances must tend toward the above model (integrated moving average process) as sampling interval gets larger.

For the above model,

$$\begin{aligned} y(k|k-1) &= \left(1 - \frac{1-q^{-1}}{1-\alpha q^{-1}}\right) y(k) \\ &= \frac{(1-\alpha)q^{-1}}{1-\alpha q^{-1}} y(k) \\ &= (1-\alpha)[y(k-1) + \alpha y(k-2) + \alpha^2 y(k-3) \cdots\cdots \end{aligned}$$

This is the EWMA. Hence, EWMA is thought to provide more effcient monitoring under the postulated model due to its extrapolation capability.

In addition, the prediction error (or whitened output) becomes

$$\varepsilon(k) = \frac{\Delta y(k)}{1 - \alpha q^{-1}} \tag{1.40}$$

which can be written as

$$\varepsilon(k) = \alpha\varepsilon(k-1) + \Delta y(k) \tag{1.41}$$

This is EWMA for the differenced output.

1.4.7 MULTIVARIABLE TIME-SERIES MODEL

Getting Rid of Both the Spatial and Temporal Correlations
A natural next step is to consider the time correlation in the multivariate statistical monitoring context.

MIMO Time Series Model?
One option is to fit a multivariable time series model to the data. For instance, MIMO ARMA model looks like

$$y(k) = A_1 y(k-1) + \cdots + A_n y(k-n) + \varepsilon(k) + C_1\varepsilon(k-1) + \cdots + C_n\varepsilon(k-n) \tag{1.42}$$

Once a model of the above form becomes available, one can then compute the prediction error as before (which is a white sequence) and apply the

chi-square monitoring.

The trouble is that multivariable time-series models are notoriously difficult to construct from data. It requires a special parametrization of the coefficient matrices and the resulting regression problem is nonconvex with many local minima.

State-Space Model?

A much better option is to fit the following state-space model instead:

$$\begin{aligned} x(k+1) &= Ax(k) + K\varepsilon(k) \\ y(k) &= Cx(k) + \varepsilon(k) \end{aligned} \tag{1.43}$$

Such models may be developed from y data using one of the following options:

- computation of the autocorrelation function followed by factorization and model reduction.
- state-space realization using (modified) *subspace identification* techniques. The obtained model can be further refined using the prediction error minimization.

Once a model of the above form becomes available, $x(k)$ in the model can be updated recursively. The prediction error $\varepsilon(k) = y(k) - Cx(k)$ is a time-wise independent sequence and can be monitored using the chi-square statistics as before. The two-tier approach based on the principal component analysis should be utilized in this context as well.

Incorporating Deterministic Inputs

If there are deterministic inputs, one can include them in the model as before.

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) + K\varepsilon(k) \\ y(k) &= Cx(k) + \varepsilon(k) \end{aligned} \tag{1.44}$$

Model for Nonstationary Sequences

Finally, in the case that the outputs exhibit nonstationary, drifting-type behaviors, the data can be differenced before the model fitting and this results in a model of the form

$$\begin{aligned} x(k+1) &= Ax(k) + B\Delta u(k) + K\varepsilon(k) \\ \Delta y(k) &= Cx(k) + \varepsilon(k) \end{aligned} \tag{1.45}$$

1.5 REGRESSION

1.5.1 PROBLEM DEFINITION

We have two vector variables $x \in \mathcal{R}^n$ and $y \in \mathcal{R}^p$ that are *correlated*. We have a data set consisting of N data points, $\{(x(i), y(i)), i = 1, \cdots, N\}$. We assume that x and y are both mean-cetered (x and y represent the deviation variables from the respective means). Now, using the data, we wish to construct a prediction model

$$\hat{y} = f(x) \tag{1.46}$$

which can be used to predict y given a fresh data point for x.

Example

- In a distillation column, relate the tray temperatures to the end-point compositions. In this case $x = [T(1), T(2), \cdots, T(n)]^T$ and $y = [x_D, x_B]^T$.
- In a polymer reactor, relate the temperature and concentration (trends) to the average molecular weight, polydispersity, melt index, etc. of the product.

- In a pulp digester, relate the liquor concentrations and temperature to the wood composition (e.g., Kappa Number) of the product.

1.5.2 THE METHOD OF LEAST SQUARES

What Is It?

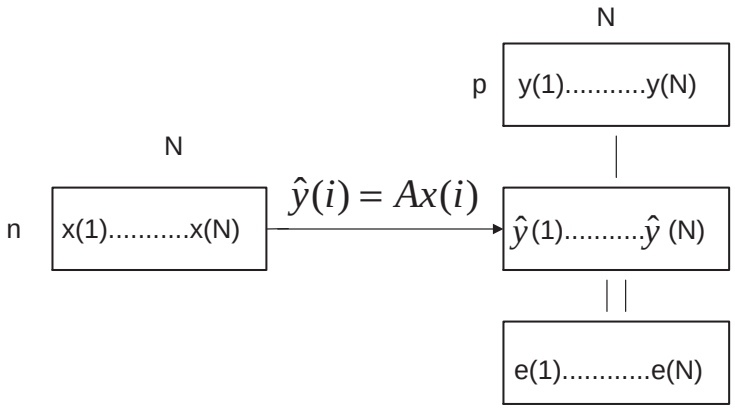
The most widely used is the method of least squares. The least squares method is particularly powerful when one wishes to build a *linear* prediction model of the form

$$\hat{y} = Ax \tag{1.47}$$

With N data points, we can write

$$\underbrace{\begin{bmatrix} y(1) & \cdots & y(N) \end{bmatrix}}_{Y} = A \underbrace{\begin{bmatrix} x(1) & \cdots & x(N) \end{bmatrix}}_{X} + \underbrace{\begin{bmatrix} e(1) & \cdots & e(N) \end{bmatrix}}_{E} \tag{1.48}$$

The last term represents the prediction error (for the prediction model $\hat{y} = Ax$) for the N available data points.



A reasonable criterion for using A is

$$\min_{A} \{ \sum_{i=1}^{N} e^T(i) e(i) = \| Y - AX \|_f^2 \} \tag{1.49}$$

The solution to the above is

$$A = YX^T(XX^T)^{-1} \tag{1.50}$$

Statistical Interpretation

One can develop the least squares solution from the following statistical argument. Suppose the underlying system (from which the N data set was generated) is

$$y = Ax + \varepsilon \tag{1.51}$$

where ε is a zero-mean, Gaussian random variable vector (covering for the noise and other randomness in the relationship between x and y). Assume also that x is a Gaussian vector. Then, y is also Gaussian due to the linearity. Then,

$$E\{y|x\} = \bar{y} + \text{cov}\{y,x\}\text{cov}^{-1}\{x,x\}(x - \bar{x}) \tag{1.52}$$

Since x and y are both mean-centered variables, $\bar{x} = 0$ and $\bar{y} = 0$. We now approximate the covariances using N data points available to us.

$$\text{cov}(y,x) \approx R_{yx} = \frac{1}{N}\sum_{i=1}^{N} y(i)x^T(i) \tag{1.53}$$

$$\text{cov}(y,x) \approx R_x = \frac{1}{N}\sum_{i=1}^{N} x(i)x^T(i) \tag{1.54}$$

Hence,

$$\begin{aligned} E\{y|x\} \approx \hat{y} &= \underbrace{\left(\frac{1}{N}\sum_{i=1}^{N} y(i)x^T(i)\right)\left(\frac{1}{N}\sum_{i=1}^{N} x(i)x^T(i)\right)^{-1}}_{A} x \\ &= \tfrac{1}{N}YX^T(\tfrac{1}{N}XX^T)^{-1}x \end{aligned} \tag{1.55}$$

Note that the above is the same as the predictor that results from the method of least squares.

1.5.3 LIMITATIONS OF LEAST SQUARES

Possibility of Ill-Conditioning

Recall the least squares solution

$$\begin{aligned} \hat{y} &= YX^T(XX^T)^{-1}x \\ &= R_{yx}R_x^{-1}x \end{aligned} \tag{1.56}$$

Since R_x is a symmetric, positive (semi)-definite matrix, it has the decomposition in the form of

$$R_x^{-1} = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_1^2} & & \\ & \ddots & \\ & & \frac{1}{\sigma_n^2} \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_n^T \end{bmatrix} \tag{1.57}$$

In the case that the x data are highly correlated, $\sigma_1 \gg \sigma_n$ and some of σ's will be very small (in a relative sense).

Implication of Ill-Conditioned Information Matrix

This has the following implications.

- *Possibility of Artificially High Gains Due to Poor Signal to Noise Ratio*
  Note that R_{yx} and R_x are only approximations of the covariance matrices based on N data points. Due to the error in the data, they both contain errors. When $\frac{1}{\sigma_i^2}$'s are large, errors in R_{yx} can get amplified greatly leading to a bad predictor (*e.g.*, a predictor with artificially high gains).

- *Sensitivity to Outliers and Noise*
  Also, even if the covariance matrices were estimated perfectly, the prediction can still be vulnerable to errors in the x data due to the high gain.

- *Statistical Viewpoint*

R_x (actually XX^T) is called information matrix. σ_i represents the amount of information contained in the data X for a particular linear combination of x (given by $v_i^T x$). Hence, small σ_i means small amount of information. Naturally, extracting the correlation between $v_i^T x$ and y from the very small amount of data can lead to trouble.

Examples

CONSIDER A TWO-DIMENSIONAL CASE WITH AN ILL-CONDITIONED INFORMATION MATRIX. GRAPHICALLY ILLUSTRATE THE DATA DISTRIBUTION AND HOW IT RELATES TO THE SVD, RESULTING ESTIMATE, etc.

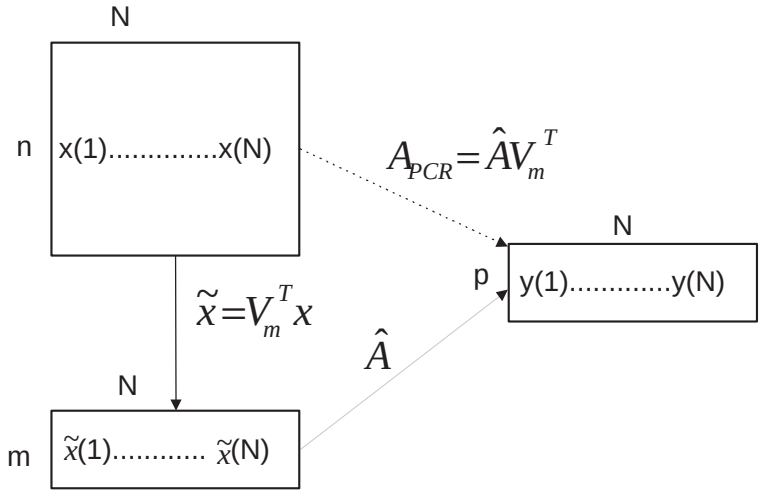
1.5.4 PRINCIPAL COMPONENT REGRESSION

Main Idea

Partition the decomposition of the matrix R_x as

$$\left[\begin{array}{ccc|ccc} v_1 & \cdots & v_m & v_{m+1} & \cdots & v_n \end{array}\right] \left[\begin{array}{ccc|ccc} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_m & & & \\ \hline & & & \sigma_{m+1} & & \\ & & & & \ddots & \\ & & & & & \sigma_n \end{array}\right] \left[\begin{array}{c} v_1^T \\ \vdots \\ v_m^T \\ \hline v_{m+1}^T \\ \vdots \\ v_n^T \end{array}\right] \tag{1.58}$$

The main idea is to project the data down to the reduced dimensional space defined by $v_1, \cdots, v_m$ (which also represents the space for which a large amount of data are available). This is illustrated graphically as follows:



We can write the projection as

$$\tilde{x} \triangleq V_m^T x = \begin{bmatrix} v_1^T \\ \vdots \\ v_m^T \end{bmatrix} x \tag{1.59}$$

$\tilde{x}$ represents the principal components of x. Note that, in the case that x is of very high dimension, it is likely that $\dim\{\tilde{x}\} \ll \dim\{x\}$. We can write the least squares estimator as

$$\begin{aligned} \hat{y} &= \underbrace{Y\tilde{X}^T(\tilde{X}\tilde{X}^T)^{-1}}_{\tilde{A}}\tilde{x} \\ &= \underbrace{\tilde{A}V_m^T}_{A_{PCR}}x \end{aligned} \tag{1.60}$$

This is called *principal component regression.*

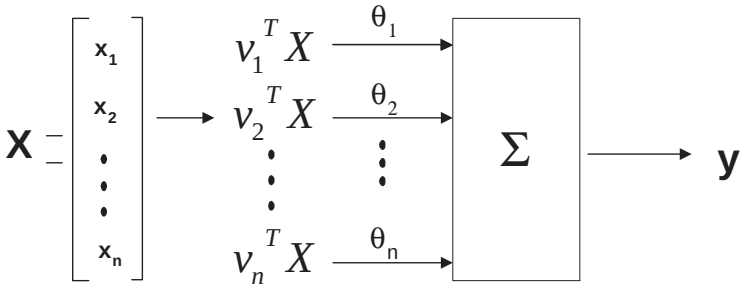
Statistical Viewpoint

In a statistical sense, it can be interpreted as accepting *bias* for reduced *variance*. We are *a priori* setting the correlation between

$v_i^T x, i = m+1, \cdots, n$ and y to be zero, i.e.,

$$y = \theta_1 \underbrace{v_1^T x}_{\tilde{x}_1} + \cdots + \theta_m \underbrace{v_m^T x}_{\tilde{x}_m} + 0 \times \underbrace{v_{m+1}^T x}_{\tilde{x}_{m+1}} + \cdots + 0 \times \underbrace{v_n^T x}_{\tilde{x}_n}$$

since computing the correlation based on data can introduce substantial variances which are thought to be much more harmful to estimation than the bias.



Example

TAKE THE PREVIOUS EXAMPLE, DO THE PRINCIPAL COMPONENT REGRESSION AND SHOW THE VARIANCE VS. BIAS TRADE-OFF.

1.5.5 PARTIAL LEAST SQUARES (PLS)

Main Idea

PLS is similar to PCR in that they are both biased regressions or subspace regression. The difference is that, in PLS, the subspace (consisting of m directions) of the regressor space is chosen to maximize XY^TYX^T rather than XX^T. In other words, in choosing the m directions, one looks at

- not only how much a certain modes contributes to the X data,
- but also how much it is correlated with the Y data.

In this sense, it can be thought as a middle ground between the PCR and the regular least squares.

Procedure

The PLS procedure can be explained as follows:

1. Set $i = 1$. $X_1 = X$ and $Y_1 = Y$.
2. Choose the principal direction v_i for $X_i^T Y_i Y_i^T X_i$.
3. $v_i^T X_i = \tilde{X}_i$
4. Compute the LS prediction of Y_i based on $\tilde{X}_i$.

$$\hat{Y}_i = Y_i \tilde{X}_i^T \frac{1}{(\tilde{X}_i \tilde{X}_i^T)} \tilde{X}_i \tag{1.61}$$

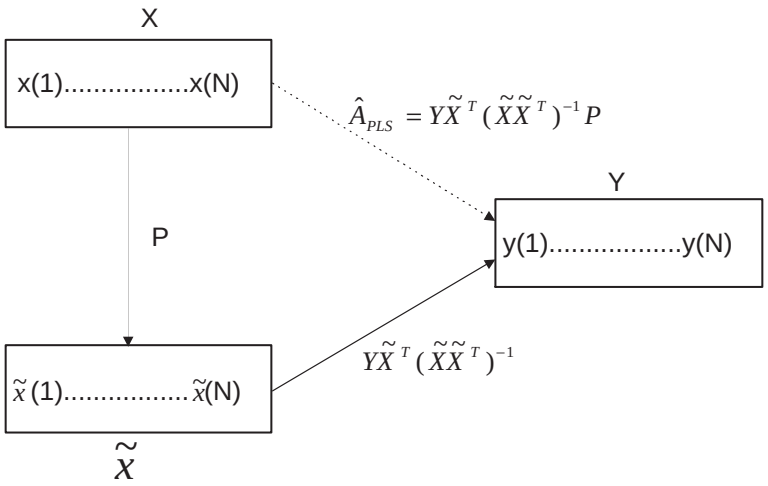
5. Compute the residuals

$$\begin{aligned} Y_{i+1} &= Y_i - \hat{Y}_i \\ X_{i+1} &= X_i - v_i \tilde{x}_i \end{aligned} \tag{1.62}$$

6. If the residual Y_{i+1} is sufficiently small, stop. If not, set $i = i + 1$ and go back to Step 2.

From the above, with m iteration, the m-dimensional regression space is defined. Create a data matrix

$$\tilde{X} = \begin{bmatrix} \tilde{X}_1 \\ \vdots \\ \tilde{X}_m \end{bmatrix}$$



$\tilde{X}$ can be expressed as a linear projection of X:

$$\tilde{X} = PX \tag{1.63}$$

where $P \in \mathcal{R}^{m \times n}$. Then, the PLS predictor can be written as

$$\begin{aligned} \hat{y} &= Y\tilde{X}^T(\tilde{X}\tilde{X}^T)^{-1}\tilde{x} \\ &= \underbrace{Y\tilde{X}^T(\tilde{X}\tilde{X}^T)^{-1}P}_{A_{PLS}}x \end{aligned} \tag{1.64}$$

The above is not the most efficient algorithm from a computational standpoint. The most widely used is the NIPALS (Nonlinear Iterative Partial Least Squares) algorithm described in (Geladi and Kowalski, Analytica Chimica Acta, 1986)

1.5.6 NONLINEAR EXTENSIONS

Regression needs not be confined to just linear relationships. More generally, one can search for a prediction model of the form $\hat{y} = f(x)$ where f can be a nonlinear function.

Finite Dimensional Parameterization

To reduce the problem to a parameter estimation, one defines a search set with a finite dimensional parameterization for the nonlinear function. The search set can take on different forms.

- **Functional Expansion**

$$\hat{y} = \sum_{i=1}^{n} c_i \phi_i(x) \tag{1.65}$$

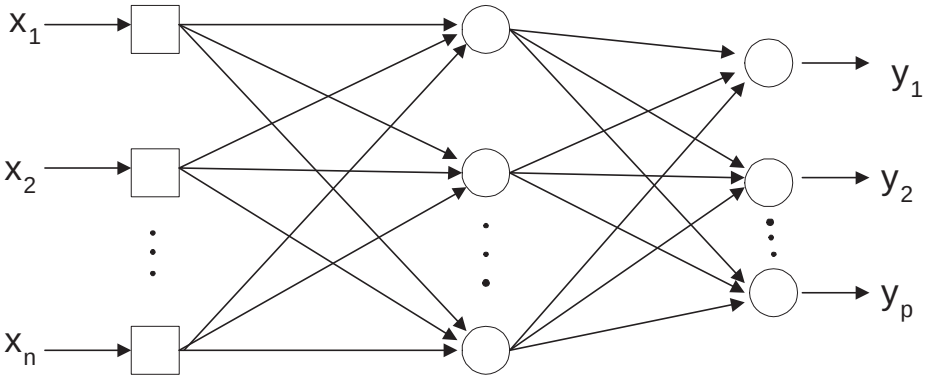
or

$$\hat{y} = \sum_{i=1}^{n} \phi_i(x, c_i) \tag{1.66}$$

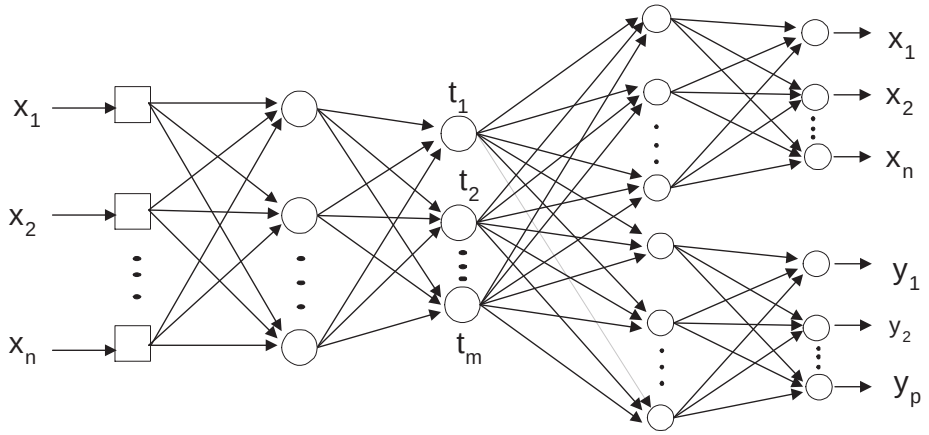
$\{\phi_i(x), i = 1, \cdots, n\}$ are basis functions (polynomials, sinusoids, wavelets, Gaussian functions, etc.). The problem is reduced to finding c_i. The former leads to a linear regression problem while the latter to a nonlinear problem. The order can be determined on an iterative basis, that is, by examining the prediction error as more and more terms are introduced.

- **Network Based Approach**

For instance, shown below is the so called Artificial Neural Network (ANN) inspired by biological neural systems. The parameters are the various weights which must be selected on the basis of the available data. This is referred to as the *learning* in the ANN parlance. The usual criterion is again the least squares or its extensions.



Nonlinear PLS



1.5.7 EXTENSIONS TO THE DYNAMIC CASE

Suppose x and y have dynamic correlations:

$$y(k) = f(x(k), x(k-1), \cdots, \cdots) \tag{1.67}$$

Different structures can be envisioned:

- *Time Series:* construct a predictor of form

$$\begin{aligned} \hat{y}(k) &= a_1\hat{y}(k-1) + \cdots + a_n\hat{y}(k-n) \\ &\quad + b_0x(k) + b_1x(k-1) + \cdots + b_mx(k-n) \\ y(k) &= \hat{y}(k) + \varepsilon(k) \end{aligned} \tag{1.68}$$

  $a_1, \cdots, a_n, b_0, \cdots, b_m$ can be found to minimize the prediction error using the available data. Note that, since we don't have data for $\hat{y}(k-1), \cdots, \hat{y}(k-n)$, and they depend on the choice of the parameters, this is a nonlinear regression problem. Therefore, it is pretty much limited to SISO problems.

- *State-Space Model:* For MIMO systems, use Subspace ID to create

$$\begin{aligned} z(k+1) &= Az(k) + \varepsilon_1(k) \\ \begin{bmatrix} x(k) \\ y(k) \end{bmatrix} &= \begin{bmatrix} C_x \\ C_y \end{bmatrix} z(k) + \varepsilon_2(k) \end{aligned} \tag{1.69}$$

  Then, build the Kalman filter that uses the measurement x to predict y can be written as

$$\begin{aligned} \hat{z}(k) &= A\hat{z}(k-1) + K(x(k) - C_xA\hat{z}(k-1)) \\ \hat{y}(k) &= C_y\hat{z}(k) \end{aligned} \tag{1.70}$$

Chapter 2

APPLICATION AND CASE STUDIES

Outline

- SBR Semi-Batch Reactor System: Monitoring
- Batch Pulp Digester: Inferential Kappa Number Control
- Nylon 6,6 Autoclave: Monitoring & Inferential Control of Quality Variables
- Continuous Pulp Digester: Inferential Kappa Number Control

2.1 PCA MONITORING OF AN SBR SEMI-BATCH REACTOR SYSTEM

2.1.1 INTRODUCTION

Background

- In operating batch reaction systems, certain *abnormalities* (e.g., increased feed impurity level, catalyst poisonging, instrumentation malfunctioning) develop that eventually throw the quality completely off spec.
- It is desirable to catch these incipient faults quickly so that the problem can be rectified.
- It is desirable not to rely on lab measurements for this purpose since this will introduce significant delays.

Key Idea

- Use more easily measured process variable trends to classify between normal batches and abnormal batches.
- The key problem is to extract out the key identifying features (*finger prints*) from trajectories of large amount of variables.

Appication

- An SBR Polymerization Reactor.

2.1.2 PROBLEM DESCCRIPTION

Process / Problem Characteristics

- Reaction:

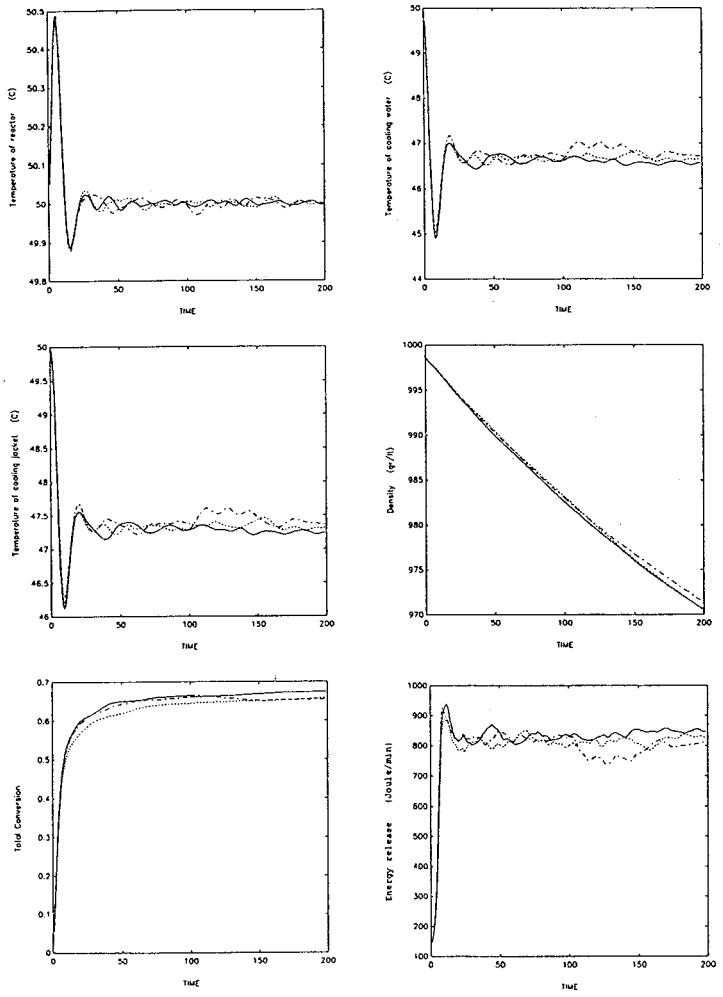
$$\text{Styrene} + \text{Butadiene} \longrightarrow^{\text{polymerization}} \text{Latex Rubber}$$

- Emulsion Polymerization
- The reactor is initially charged with seed SBR particles, initiator, chain-transfer agent, emulsifier, a small amount of styrene and butadiene monomers.
- Batch duration is 1000 minutes.
- The following measurements are available with 5 minute interval:
  - flow rates of styrene
  - flow rates of butadiene
  - temperature of feed
  - temperature of reactor
  - temperature of cooling water
  - temperature of reactor jacket
  - density of latex in the reactor
  - total conversion (an estimate)
  - instantaneous rate of energy release (an estimate)

Available Data

- 50 batch runs with typical random variations in base case conditions (such as initial charge of seed latex, amount of chain transfer agent and level of impurities).

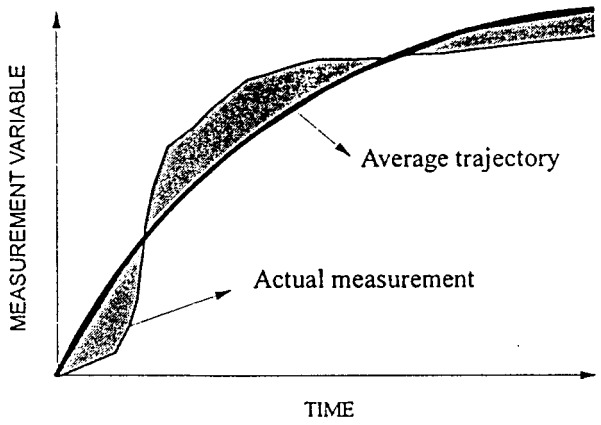
- Two additional batches with "unusual disturbances."
  - impurity of 30% above that of the base case was introduced in the butadiene feed at the beginning of the batch.
  - impurity of 50% above that of the base case was introduced in the butadiene feed at the halfway mark.



2.1.3 RESULTS

End-Of-Batch Principal Component Analysis

- Establish the mean trajetory for each variable and compute the deviation trajectory.



- Normalize each variable with its variance.
- Perform "lifting", that is, stack all the trajectories into a common vector to obtain a single vector $\mathcal{Y}$ for each batch. Then, form a matrix Y by aligning $\mathcal{Y}$ for the entire 50 batches.

  Note the dimension of $\mathcal{Y}$ is 9×200. Clearly there are only a few modes of variations in this vector.

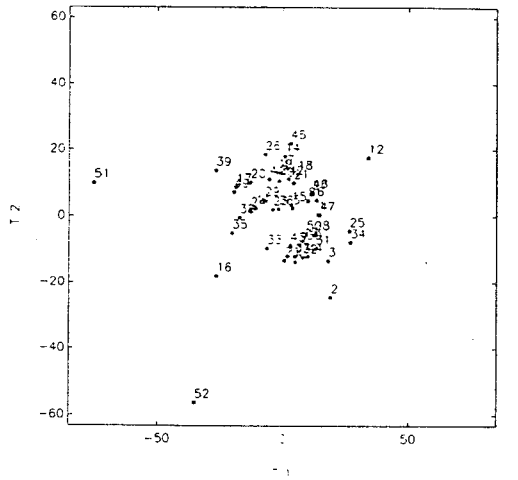
- Determine the principal component directions (eigenvectors of Y with significant eigenvalues). Three components were judged to be sufficient.

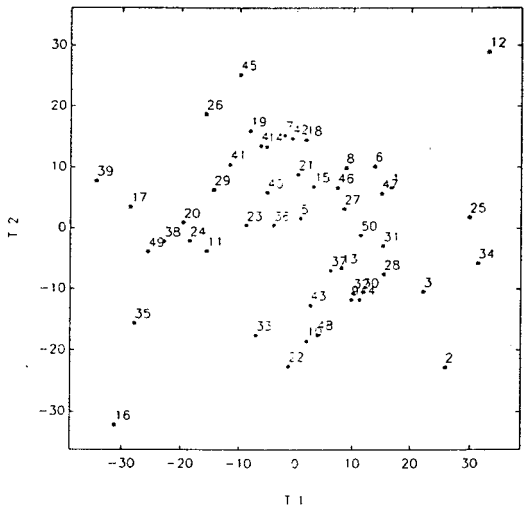
$$Y = \left[\begin{array}{ccc|ccc} v_1 & v_2 & v_3 & v_4 & \cdots & v_{1800} \end{array}\right] \left[\begin{array}{ccc|ccc} \sigma_1 & & & & & \\ & \sigma_2 & & & & \\ & & \sigma_3 & & & \\ \hline & & & \sigma_4 & & \\ & & & & \ddots & \\ & & & & & \sigma_{1800} \end{array}\right] \left[\begin{array}{c} v_1^T \\ v_2^T \\ v_3^T \\ \hline v_4^T \\ \vdots \\ v_{1800}^T \end{array}\right]$$

- Compute the principal component score variables for each batch:

$$t_i(j) = v_i^T \mathcal{Y}(j), \qquad i = 1, \cdots, 3 \quad j = 1, \cdots, 50$$

The first two P.C. scores for the 50 batches and the two bad batches are plotted below:





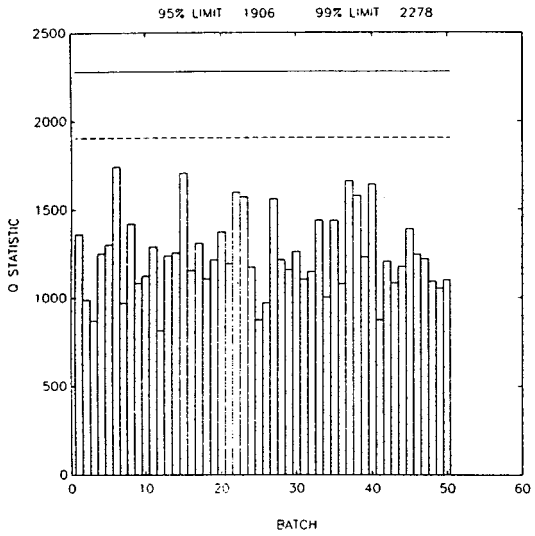
Compute the covariance matrix (diagonal) R_t for the P.C.'s. Establish the 95% and 99% confidence limits (ellipses) for the P.C.'s.

One can also use Hotelling Statistics:

$$D = t^T R_t^{-1} t \frac{N(N-m)}{m(N^2-1)} \sim F_{m,N-m}$$

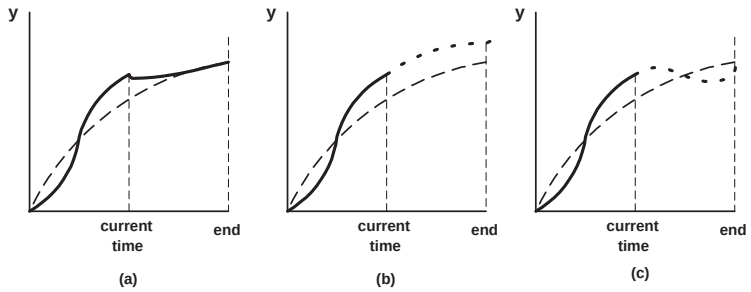
Here $t = [t_1,\ t_2,\ t_3]^T$ and $N = 50$ and $m = 3$.

- Compute the residuals and establish the 95% and 99% confidence limits for the square sum (assuming normality of the underlying distribution). The SPE (sum of the squares of the residuals) for each batch is plotted against the confidence limits:



During-Batch Principal Component Analysis

- The main issue in applying the PC monitoring during a batch is what to do with the missing future data.



Handling missing measurement

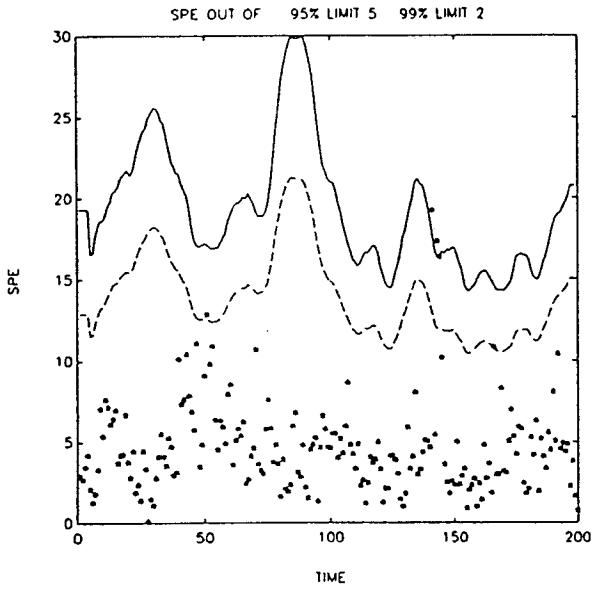
Options are:

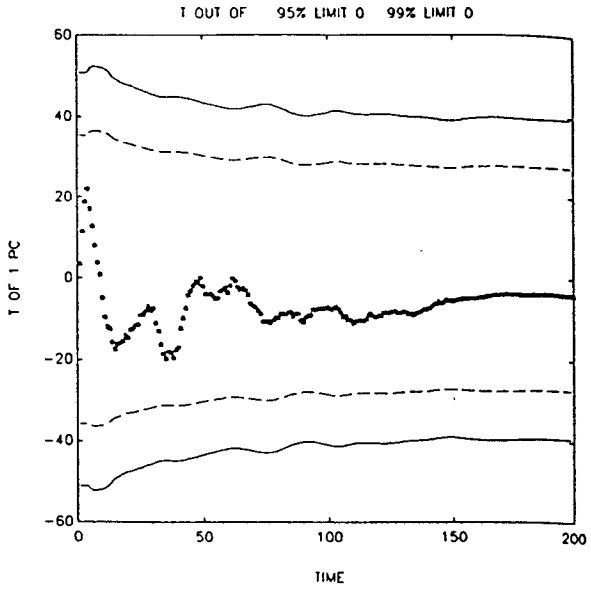
- Assume for all the variables that the future deviation will be zero.
- Assume for each variable that the current level of deviation will continue until the end of batch.
- Use statistical correlation to estimate the future deviation.

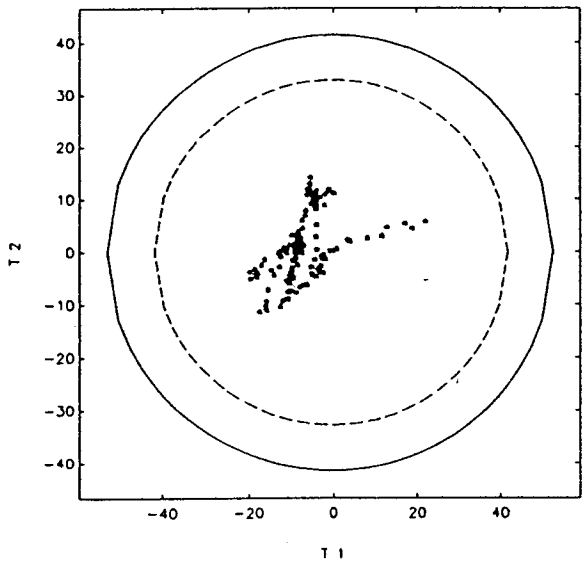
We will denote the lifted vector at time t with missing future measurements filled in as $\hat{\mathcal{Y}}_t(j)$, where t and j denote the time and batch index.

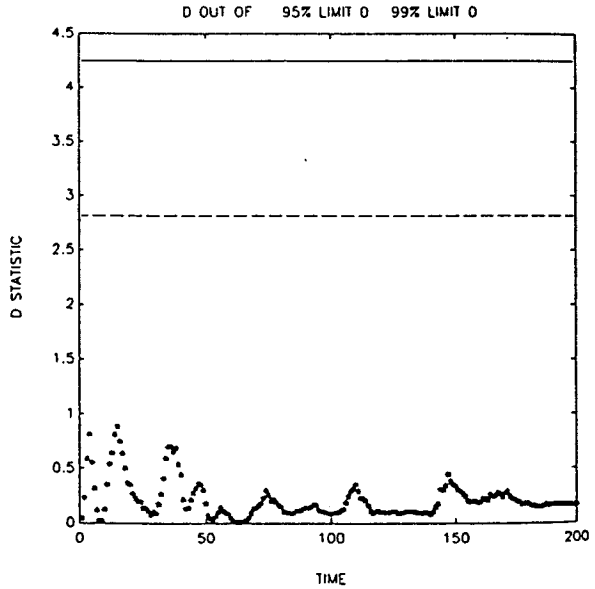
- For each time step, the confidence limits for the SPE and P.C.'s can be established.
- Now, at each time step for each batch, compute the P.C.s and SPE and compare against the confidence intervals.

Good Batch

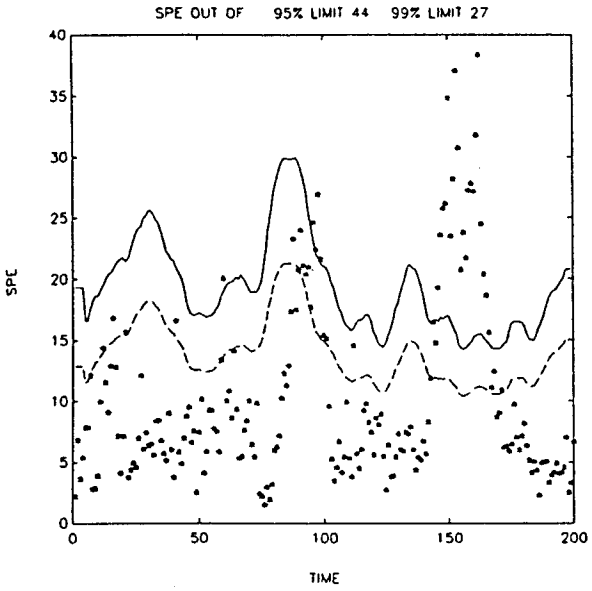


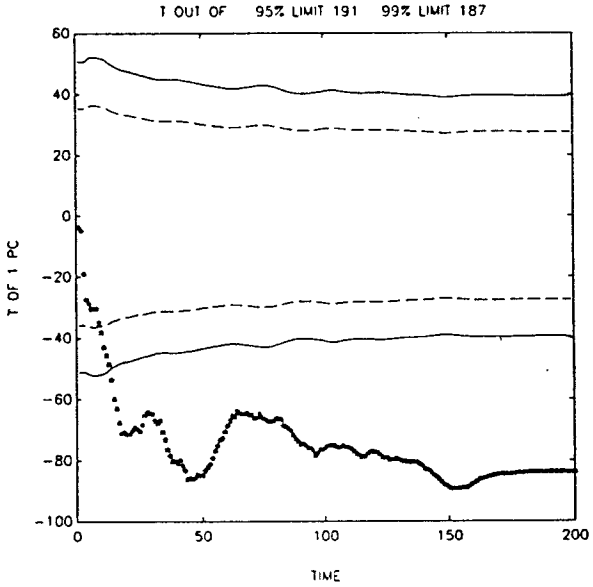


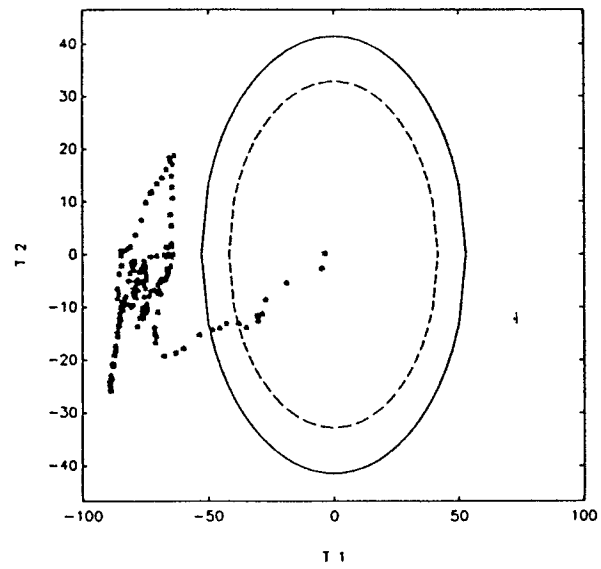


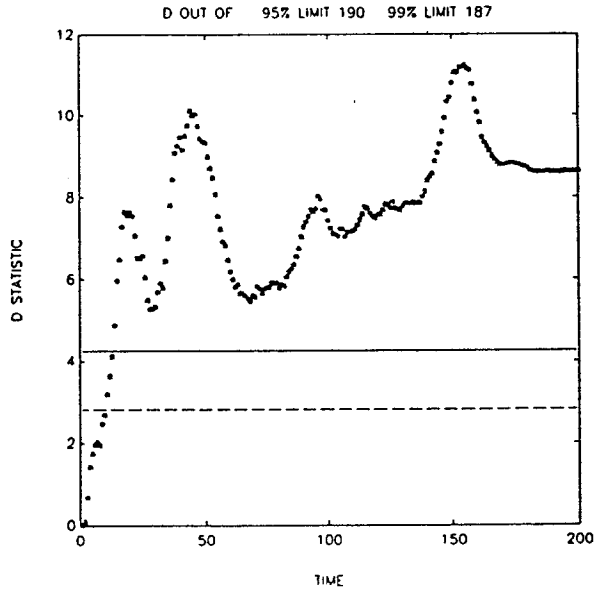


Bad Batch I

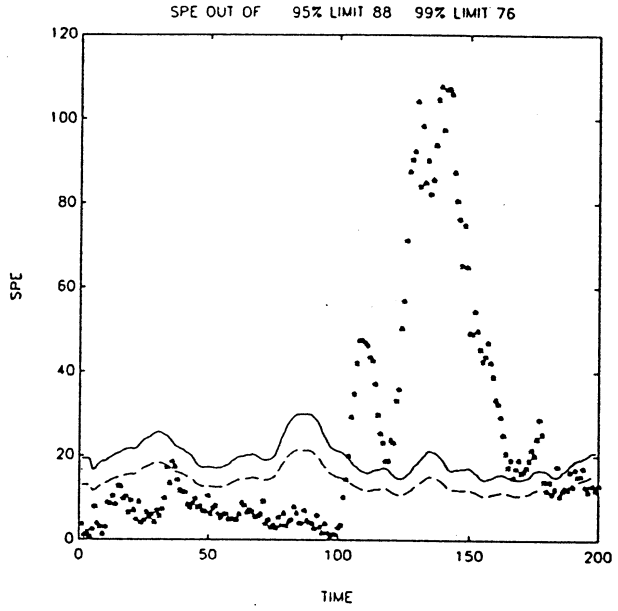


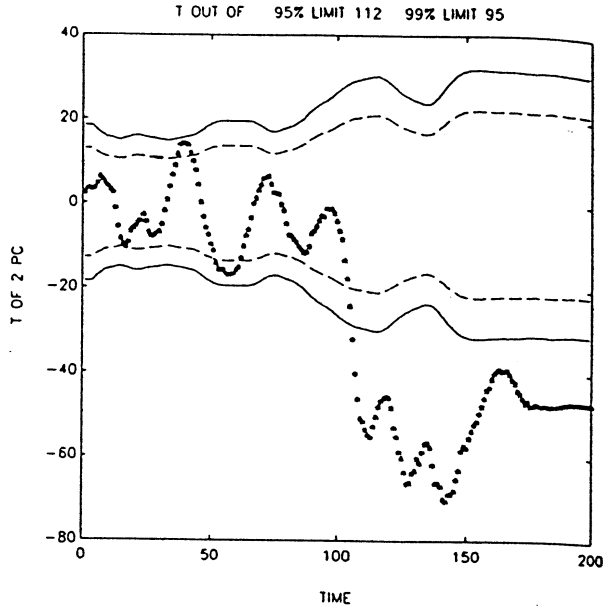


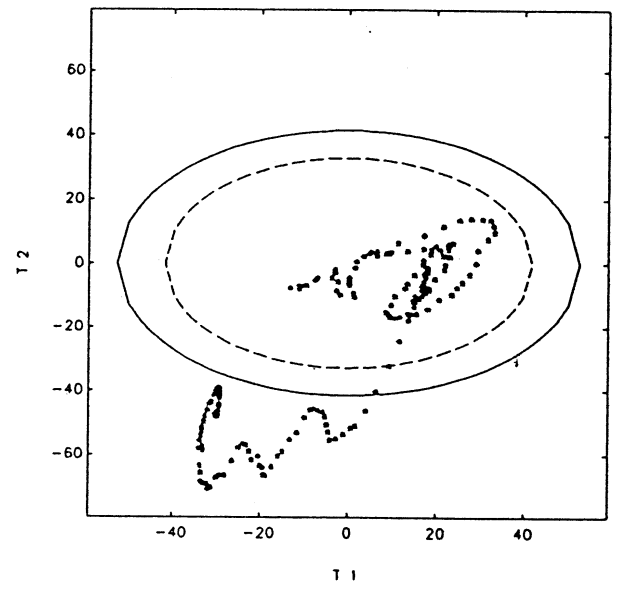


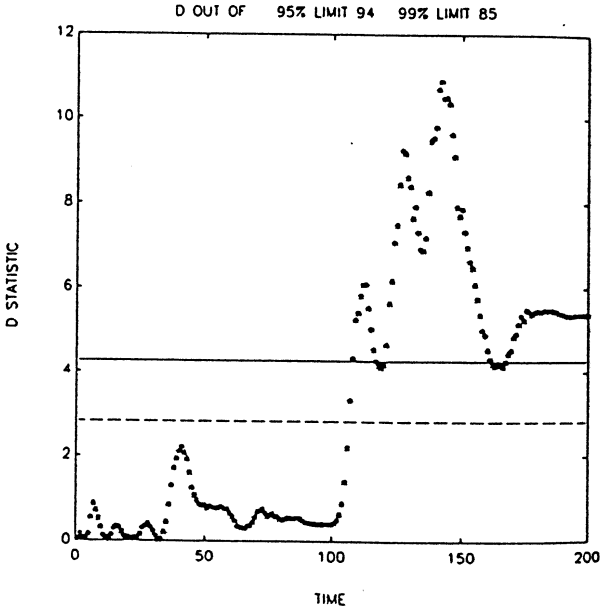


Bad Batch II









2.2 DATA-BASED INFERENTIAL QUALITY CONTROL IN BATCH REACTORS

2.2.1 INTRODUCTION

Background

- Large quality variances are often due to
  - machine (equipment, instrumentation) errors for which the dinstinction betwewn failures and nonfailures are clear
  - feed variations and operating condition variations for which the dividing line between failures and nonfaillures is often blurred.
- The latter disturbances tend to fluctuate quite a bit from batch to batch and are usually not removable at source.
- For these disturbances, on-line prediction and control are desired rather than statistical monitoring followed by diagnosis (since these cannot be categorized as as Pareto's glitches).

Key Idea

- Capture the statistical correlation between easily measured process variable trajectories and final quality variables through regression.
- Use the regression model for on-line prediction and control (through manipulation of operating parameters) of final quality variables.

Application

- Batch Pulp Digester
- Nylon 6,6 Autoclave

2.2.2 CASE STUDY IN DETAILS

See the attached!

2.3 INFERENTIAL QUALITY CONTROL OF CONTINUOUS PULP DIGESTER

2.3.1 INTRODUCTION

Background

- In continous systems, on-line quality measurements are often (1) very difficult, (2) very expensive, and/or (3) unavailable.
- Lab measurements introduce large delays, making tight control impossible (high-frequency errors are pretty much left uncontrolled).
- There is significant incentive to reduce the variability by increasing the bandwidth of control.

Key Idea

- Relate more easily measured process variables to quality variables *dynamically* through data regression.
- Use the regression model for on-line prediction and control of quality variables.
  Elimination of delays $\rightarrow$ More efficient prediction of quality variables $\rightarrow$ tighter control.

Appication

- A continuous pulp digester.

2.3.2 CASE STUDY IN DETAILS

See attached!