QUALITY AND RELIABILITY ENGINEERING INTERNATIONAL, VOL. 9, 89–93 (1993)

# SPC OF A NEAR ZERO-DEFECT PROCESS SUBJECT TO RANDOM SHOCKS

M. XIE AND T. N. GOH

*Department of Industrial and Systems Engineering, National University of Singapore, Kent Ridge, Singapore 0511, Republic of Singapore*

SUMMARY

In this paper we study the problem of monitoring and control of a type of process in which long series with no non-conformities are observed together with occasional samples containing a large number of non-conformities. We call this a near zero-defect process subject to random shocks. Such processes occur often in practice, and a model is proposed for the identification of real non-random variations of process characteristics. Based on the statistical analysis carried out for this model, a procedure for decision-making in the control of this type of process is suggested, and analysis of some actual cases presented.

KEY WORDS Zero defect Random shocks Statistical process control Non-conformity charts

## 1. BACKGROUND

Statistical process control based on control charts has been in use in industry for decades.[1–3] In a near 'zero-defect' production environment, such as those found in the electronics industry conventional control chart users tend to face the problem of long series of samples with no non-conformity between occasional ones containing non-conformities. This is not a phenomenon addressed by Shewhart control charts. Modification and improvement of existing process control procedures thus have to be developed for more effective decision making; see, for example, References 4 to 7.

As an example, data on read-write errors discovered in a computer hard disk can exhibit a string of zeros before a large count of errors suddenly appears. Table I exhibits a set of actual data.

Table I. A typical data set of non-conformities

Data set 1 (number of non-conformities; read from left to right):

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 1 | 4 | 0 | 75 | 0 | 0 | 0 | 0 | 0 |
| 0 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 3 | 0 | 0 | 0 | 0 | 0 | 9 | 0 | 0 |
| 1 | 3 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 |
| 0 | 5 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 15 | 0 | 75 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 6 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 9 | 0 | 0 | 0 | 2 | 1 | 0 | 0 | 0 | |
| 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |

The number of errors, or non-conformities, is typically assessed by the u-chart. The control limit for the u-chart is determined to be 5 for data set 1. In this case, many points fall outside the limit, and usually no obvious causes can be found for such 'out-of-control' points. There are thus many false alarms if the conventional u-chart is used: in this case, even when the exact probability limit is used, the upper control limit is 6 with the same consequential difficulties in decision-making.

In this paper, we propose a two-stage model for this particular kind of process. It is understood that the individual product units are usually perfect, but there exists a random mechanism, here called shocks, which leads to a number of non-conformities in the sample whenever it is activated. The shocks occur randomly with a low probability so that a large number of zero-non conformity samples are observed; however, when a sample is affected by a shock, it can contain any number of non-conformities. Realization of this kind of process may give us data similar to that in Table I.

We call the model one of a zero-non-conformity process subject to random shocks. The model is applicable in many practical situations in which some kind of unidentified and uncontrollable mechanism takes place to cause non-conformities in an otherwise near-perfect process. The proposed model is useful in developing a procedure for determining whether an out-of-control situation exists.

Section 2 presents the main idea of the model and Section 3 provides a basic analysis for it. In Section 4 we derive the maximum likelihood esti-

0748–8017/93/020089–05$07.50


*Received 20 June 1992*
*Revised 10 November 1992*

mates for the model parameters using available data, which can be used to determine the distributional form for the construction of a modified control chart as well as a recommended decision-making process studied in Section 5. Finally, we give some practical examples in which the approach has been used.

## 2. MATHEMATICIAL MODEL

The proposed two-stage model for the analysis of non-conformity data with many zeros is described below. In this model, each sample is treated as a unit and each unit contains a number of discrete components. For each unit, we assume that

(a) there is a probability, $p$, that the unit is a non-conforming (defective) one containing some non-conformities, i.e. non-conforming components. Therefore, $1-p$ is the probability that the unit is perfect
(b) in a non-conforming unit, the number of non-conformities in this unit follows a certain probability distribution.

In an ideal situation, only perfect units are observed. However, a unit can be affected by a random mechanism that introduces non-conformities. If such a mechanism (which may or may not be eventually controllable) occurs, there will be some non-conformities in the unit.

As another illustration, suppose a transportation truck transports a large number of small components in cartons and, on arrival at the destination, the numbers of damaged components are counted. We assume that a carton will contain damaged components only if the carton itself is damaged during transportation, in which case a number of components may be damaged. The number of components in the carton that are damaged is then a random quantity that follows a distribution, independent of the probability that a carton is damaged during transportation. Since it is not often that a carton is damaged, many zero-defect or zero-non-conformity cartons are counted. However, when a carton contains non-conformities, it may contain any number of them, i.e. damaged components.

Note that there is a possibility that a non-conforming unit contains no non-conformities. Thus, in the above example, there is a probability that a carton is in some way damaged without affecting any component inside. Although there is no need to distinguish between a non-conforming unit with no non-conformities and a perfect unit, it may still be of interest to estimate the probability of the occurrence of the random mechanism, $p$.

## 3. DISTRIBUTION OF THE NUMBER OF NON-CONFORMITIES IN A SAMPLED UNIT

By the law of total probability, we can calculate the probability of a certain number of non-conformities in a random unit as follows:

$$\begin{aligned} p_0 &= P(\text{no non-conformities in the sample}) \\ &= P(\text{unit is conforming}) \\ &\quad + P(\text{unit is non-conforming}) \\ &\quad\ P(\text{no non-conformities in} \\ &\quad\ \text{a non-conforming unit}) \\ &= (1-p) + pP(\text{no non-conformities in} \\ &\quad\ \text{a non-conforming unit}) \end{aligned} \tag{1}$$

and, for $k \geq 1$,

$$\begin{aligned} p_k &= P(k \text{ non-conformities in the sample}) \\ &= P(\text{unit is non-conforming}) \\ &\quad\ P(k \text{ non-conformities in} \\ &\quad\ \text{a non-conforming unit}) \\ &= pP(k \text{ non-conformities in} \\ &\quad\ \text{a non-conforming unit}) \end{aligned} \tag{2}$$

From empirical data, we can estimate $p$ and other parameters in the distribution of $k$ non-conformities in a non-conforming unit. A decision can then be made as to whether the process is out-of-control in the sense that $p$ has become larger (more frequent occurrence of non-conforming units) or the average number of non-conformities in a non-conforming unit has become higher.

As a first step, we assume that the number of non-conformities in a non-conforming unit follows a Poisson distribution with parameter $\lambda$. The Poisson distribution, which is also a good approximation to the binomial distribution, is reasonable when the possible number of non-conformities in the unit is large. Thus,

$$P(k \text{ non-conformities in a non-conforming unit}) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{3}$$

Hence we have, from equations (1) and (2), respectively,

$$P(\text{no non-conformities in the sample}) = (1-p) + pe^{-\lambda} \tag{4}$$

and

$$\text{P}(k \text{ non-conformities in the sample}) = p\frac{\lambda^k e^{-\lambda}}{k!}, \qquad k=1,2,\ldots \tag{5}$$

The distribution of the number of non-conformities can be treated as a generalized Poisson distribution. Note that the probability of no non-conformities can be high, depending on the value of $p$, the reason behind strings of samples with no non-conformities.

## 4. MAXIMUM LIKELIHOOD ESTIMATION OF $p$ AND $\lambda$

In the model, under certain condition, $p$ is the probability of occurrence of a random shock and $\lambda$ is the average number of non-conformities in a sample when a shock manifests itself. Given a data set $\{n_k, k=0,1,2,...,n\}$ where $n_k$ is the number of samples containing $k$ non-conformities and $n$ is the total number of units in the study, we can easily estimate $p$ and $\lambda$ by the method of maximum likelihood. The likelihood function of $p$ and $\lambda$ for the data set $\{n_k, k=0,1,2,...,n\}$ is given by

$$L(p,\lambda) = \prod_{k=0}^{\infty} p_k^{n_k}$$
$$= [(1-p) + pe^{-\lambda}]^{n_0} \prod_{k=1}^{\infty} \left(p\frac{\lambda^k e^{-\lambda}}{k!}\right)^{n_k} \quad (6)$$

The logarithm of this likelihood function is

$$\ln L(p,\lambda) = n_0 \ln [(1-p) + pe^{-\lambda}] + \sum_{k=1}^{\infty} n_k \ln\left(p\frac{\lambda^k e^{-\lambda}}{k!}\right) \quad (7)$$

The partial derivatives of the log-likelihood function with respect to $p$ and $\lambda$ are

$$\frac{\partial \ln L(p,\lambda)}{\partial p} = n_0 \frac{-1 + e^{-\lambda}}{(1-p) + pe^{-\lambda}} + \sum_{k=1}^{\infty} \frac{n_k}{p} \quad (8)$$

and

$$\frac{\partial L(p,\lambda)}{\partial \lambda} = n_0 \frac{-pe^{-\lambda}}{(1-p) + pe^{-\lambda}} + \sum_{k=1}^{\infty} n_k \left(\frac{k}{\lambda} - 1\right) \quad (9)$$

The maximum likelihood estimates of $p$ and $\lambda$ can then be determined by solving the following likelihood equations

$$\frac{n_0(-1 + e^{-\lambda})}{(1-p) + pe^{-\lambda}} + \frac{n_d}{p} = 0 \quad (10)$$

and

$$\frac{-n_0 pe^{-\lambda}}{(1-p) + pe^{-\lambda}} + \frac{m}{\lambda} - n_d = 0 \quad (11)$$

where $n_d = n - n_0$ is the total number of units in which one or more non-conformities are found, and

$$m = \sum_{k=1}^{\infty} k n_k$$

is the total number of non-conformities found.

These likelihood equations have to be solved numerically. Note that only three quantities, $n_d$, $m$ and $n$, need to be supplied for the estimation.

## 5. DECISION-MAKING BASED ON THE PROCESS MODEL

We are now in a position to study decision-making procedures based on the model. If the process being studied is under control, then the occurrence of non-conformities will have a random pattern, although a large number of samples with no non-conformities are observed together with samples with many non-conformities. However, the process can deviate from its stable state and judged out-of-control. Thus as in a conventional control chart, we make use of an upper control limit (the maximum number of non-conformities in a sample) for this purpose, specifying that the probability that this will occur is less than a predetermined level (0·1%, say).

As for other statistical process control procedures, we now assume that the distributional parameters are known or estimated by analysing the existing data. Here we assume that $\lambda$ and $p$ in our model have been determined, e.g. by maximizing the likelihood equations. Let $n_u$ denote the upper control limit for a control chart based on the number of non-conformities per sample unit. Then its value can be determined by

$$P(n_u \text{ or more non-conformities in a sample}) \leq 0{\cdot}001 \quad (12)$$

It may be noted that there can be two out-of-control states which are not distinguished by the approach used here. If a large number of non-conformities are observed, this may be caused by an increase in either $\lambda$ or $p$. If $\lambda$ is increased, then it is certain that the number of non-conformities in a sample will increase. If $p$ is increased, then it will be more often that samples with non-conformities will be observed, which also lead to samples with a large number of non-conformities. Furthermore, if $\lambda$ is decreased and $p$ is increased or vice versa, then more sophisticated procedures are needed to identify yet another out-of-control state, which is not treated in this work.

With the above understanding, the decision criterion proposed for the process is equivalent to

$$\sum_{k=n_u}^{\infty} p\frac{\lambda^k e^{-\lambda}}{k!} \leq 0{\cdot}001 \quad (13)$$

or

$$(1-p) + pe^{\lambda} + \sum_{k=1}^{n_u-1} p\frac{\lambda^k e^{-\lambda}}{k!} > 0{\cdot}999 \quad (14)$$

Note that $n_u$ depends on both $p$ and $\lambda$. However, it can be easily solved because it takes only discrete values.

## 6. SOME EXAMPLES

In this section, for an appreciation of the applicability of the approach described above, we present the analysis of some actual data for read–write error testing of a certain computer hard disk. For the first two data sets, the procedure is successful. For the third data set, there is a problem of many low non-conformities together with a few large non-conformities in samples. This indicates that there is a need for a further generalization of the proposed model, e.g. by incorporating some inherent non-conformities besides the non-conformities caused by a random mechanism; this is an idea for further research.

### 6.1. *Analysis of the data set in Table I*

For data set 1, if a conventional Shewhart chart is used, then the upper control limit will be set as 5, in which case many points will fall outside. This has been shown to be impractical because of the large number of false alarms. The reason for this is that the Poisson model assumed for the conventional control limit is obviously inappropriate. The conventional u-chart has to be discarded as a result.

Using the model proposed in this paper, we have $n=208$, $m=242$ and $n_d=28$. The maximum likelihood estimates of $\lambda$ and $p$ are obtained, after solving the likelihood equations, as 8·74 and 0·135, respectively. The upper control limit is then determined to be 17 (inclusive).

Only two points are 'out-of-control' with sufficient certainty. Suppose that they are removed after the assignable causes are detected, we can, as in the traditional way, get a revised control limit by using $n=206$, $m=92$ and $n_d=26$. The revised process characteristics $\lambda$ and $p$ are 3·16 and 0·132, respectively. The upper control limit is then 8. It may then be suspected that three remaining points are due to assignable causes.

### 6.2. *Analysis of data set 2 (see Table II)*

In this case, $n=208$, $n_d=40$, $m=263$, and by numerical calculation we get the estimates of $\lambda$ and $p$ to be 6·56 and 0·193, respectively. An upper control limit with 99·9% confidence will be 14 non-conformities per sample. There are four points which fall outside this limit.

### 6.3. *Analysis of data set 3 (see Table III)*

In this case, $n=208$, $m=4676$ and $n_d=87$. The maximum likelihood estimates of $\lambda$ and $p$ are 53·75 and 0·418, respectively. The upper control limit is then determined to be 76.

For this process, there are far too many 'out-of-control' points, and the model seems to be inadequate for this data set. However, the reason may be as discussed below.

By looking at the data, we can see that there are many low (zero, one, or two numbers of non-conformities, together with quite a few large numbers of non-conformities. This may be caused by some inherent non-conformities besides the non-conformities caused by random shocks. The former is not modelled in this paper. However generalization to such cases is beyond the scope of this paper.

Another problem, which makes it impossible to

Table II. Data set 2, analysed in Section 6.2

Data set 2 (number of non-conformities; read from left to right):

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 2 | 2 | 0 | 0 | 3 | 0 | 0 | 0 | 0 |
| 0 | 1 | 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 3 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 1 | 0 | 0 | 16 | 0 | 1 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 0 | 0 | 0 | 0 | 12 | 0 | 0 | 10 | 0 | 6 |
| 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 0 |
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 6 | 15 | 0 | 89 | 0 | 0 | 6 | 0 | 0 | 1 |
| 0 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 9 |
| 9 | 0 | 0 | 3 | 0 | 0 | 0 | 3 | 0 | 0 |
| 16 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | | |

Table III. Data set 3, analysed in Section 6.3

Data set 3 (number of non-conformities; read from left to right):

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 4 | 0 | 0 | 0 | 0 | 95 | 80 | 35 | 2 | 0 |
| 0 | 2 | 0 | 16 | 0 | 0 | 0 | 2 | 0 | 0 |
| 196 | 0 | 12 | 0 | 0 | 27 | 92 | 117 | 25 | 0 |
| 0 | 92 | 3 | 0 | 0 | 21 | 188 | 0 | 0 | 0 |
| 0 | 0 | 362 | 0 | 0 | 0 | 96 | 14 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 95 | 0 | 86 | 0 |
| 0 | 0 | 0 | 0 | 83 | 0 | 221 | 6 | 0 | 0 |
| 89 | 24 | 0 | 0 | 6 | 26 | 20 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 101 | 0 | 10 | 0 |
| 0 | 0 | 0 | 0 | 0 | 4 | 68 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 6 | 2 | 92 | 0 | 20 | 1 |
| 23 | 0 | 0 | 0 | 0 | 0 | 1 | 101 | 0 | 57 |
| 98 | 0 | 0 | 0 | 2 | 3 | 312 | 5 | 0 | 0 |
| 183 | 0 | 0 | 0 | 0 | 1 | 98 | 69 | 1 | 3 |
| 0 | 0 | 0 | 0 | 1 | 0 | 71 | 0 | 0 | 0 |
| 16 | 0 | 0 | 0 | 0 | 67 | 119 | 0 | 100 | 0 |
| 21 | 1 | 0 | 0 | 77 | 86 | 111 | 2 | 0 | 0 |
| 1 | 98 | 0 | 22 | 95 | 5 | 65 | 1 | 1 | 1 |
| 0 | 0 | 0 | 0 | 95 | 0 | 7 | 0 | 98 | 0 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 0 | 0 | 1 | 6 | 92 | 4 | 11 | 0 | | |

establish an applicable model, is the instability of the process. For this data set, we can suspect that the process is far from stable, because of the consecutive large non-conformity samples. The problem of instability, however, is a common problem in applying control charts.[8]

## 7. ADVANTAGES AND POSSIBLE EXTENSIONS OF THE PROCESS MODEL

There are two situations in which we can have an out-of-control condition. First, we have a probability of observing a non-conforming unit. If this probability has assumed a value sufficiently larger than $p$, as revealed by the occurrence of more non-conforming units, then the process can be judged out-of-control. The probability of observing one unit with a large number of non-conformities is also increased accordingly.

Secondly, if $p$ is not changed, but the average number of non-conformities in a sample, $\lambda$, is changed, then we also have a higher probability of observing a unit containing a large number of non-conformities. The process can then be judged out-of-control in the sense that the random mechanism has changed.

The main advantages of the proposed model are the following:

1. The model is theoretically simple yet realistic.
2. The model leads to convenient decision-making procedures.
3. Two practical out-of-control conditions are handled at the same time.

Some possible generalizations of the proposed model are as follows:

1. In order to distinguish the two out-of-control situations, there is a need to make a bivariate control chart, one chart for the entire data and one for non-zero non-conformity.
2. The model can also be generalized by adding another Poisson random quantity for the inherent non-conformities, such as suggested in the analysis of data set 3 in Section 6.3.
3. Similar models can also be developed for other types of control charts for the number of non-conforming units.

## REFERENCES


1. E. L. Grant and R. S. Leavenworth, *Statistical Quality Control*, McGraw-Hill, New York, 1988.
2. D. C. Montgomery, *Introduction to Statistical Quality Control*, Wiley, New York, 1991.
3. A. J. Duncan, *Quality Control and Industrial Statistics*, Irwin, Homewood, 1986.
4. T. W. Calvin, 'Quality control techniques for "zero-defects"', *IEEE Trans. Components, Hybrids, and Manufacturing Technology*, **CHMT-6**, 323–328 (1983).
5. T. N. Goh, 'A control chart for very high yield processes', *Quality Assurance*, **13**(1), 18–22 (1989).
6. T. N. Goh, 'A charting technique for control of low-defective production', *International Journal of Quality and Reliability Management*, **4**(1), 53–62 (1987).
7. M. Xie and T. N. Goh, 'Some procedures for decision making in controlling high yield processes', *Quality and Reliability Engineering International*, **8**, 355–360 (1992).



*Authors' biographies:*

**M. Xie** obtained his Ph.D. in 1987 from the Division of Quality Technology at Linkoping University in Sweden. He graduated from the Royal Institute of Technology in Sweden and received an MSc in applied mathematics in 1984. In 1991 he became one of the first recipients of a newly established Lee Kuan Yew postdoctoral fellowship, the most prestigious award for researchers in Singapore, tenable at the National University of Singapore. His interests include applied statistics, system reliability analysis, software reliability and statistical process control. He has published many papers, and his book *Software Reliability Modelling* was published by World Scientific. He is a member of the American Statistical Association, the IEEE Reliability Society and the Society of Reliability Engineers.

**T. N. Goh** obtained his BE from the University of Saskatchewan and his Ph.D. from the University of Wisconsin-Madison, U.S.A. For the past two decades he has been active in teaching, research and industrial consultancy in the application of statistical methodologies for quality and productivity improvement. He is now Professor and Head of Industrial and Systems Engineering Department, National University of Singapore. He is Founder President of the Institute of Industrial Engineers, Singapore, and a senior member of a number of international professional bodies.