Effective Monitoring of Processes with Parts Per Million Defective
A Hard Problem!

Stefan H. Steiner and R. Jock MacKay

Dept. of Statistics and Actuarial Sciences, University of Waterloo, Waterloo, N2L 3G1 Canada

Abstract. Many processes have defective rates measured in parts per million (PPM). When the process yields such a high level of quality, traditional methods of process improvement, such as designed experiments, and control charts on the number of defectives or the time between defectives, are no longer effective. However, it may still be desirable to monitor and/or improve such processes. In this article, we take a critical look at attempts to apply control charts in this situation. As an alternative, we suggest that since defectives are so rare, we should carefully study any that are observed. By comparing the characteristics of the defectives to good units, both in terms of their physical dimensions and properties, and the process records from their production, we may be able to identify the key differences. Using this type of retrospective study, the goal is to identify an explanatory continuous variable or variables that can be monitored instead of the number of defectives or the time between defectives.

1 Introduction

Currently, many processes produce defectives at a rate less than 100 parts per million (PPM). We shall refer to such a process as a PPM process. For any process monitoring scheme, quickly detecting a deterioration in the defective rate is desirable. Since a PPM process produces so few defectives, looking at the number of defectives in a sample or the time between defectives does not provide much information on per unit basis. This information problem is accentuated if we are not employing 100% inspection. In this article we will address the question: "Can statistical methods be used to monitor changes in the rate of defectives in a PPM process?"

To answer this question we should first clearly define the problem. We assume that each part can be judged defective or not defective. This is a discrete environment where all non defective units are considered equivalent. The framework implies that either there is no known underlying continuous measurement or that it is not (cheaply) observable. For example, the environment does not include the situation where a part is called defective because a continuous measurement is outside of specification. If this were the case, the problem is greatly simplified by monitoring the underlying continuous

measurement. We have also excluded the possibility of using compressed limits where it is possible to classify units based on artificial specification limits that are narrower than the actual specifications (Geyer et al. 1996). In this way, the problem of PPM defectives can be transformed into one with a much larger "defective" rate that is more amenable to standard monitoring methods.

One suggested monitoring tool for this discrete environment is a control chart of the number of defectives or the time between defectives (Montgomery, 1991, Nelson, 1994, McCool and Joyner-Motley, 1998). Unfortunately, as will be shown, these previously proposed control charts do not work well for monitoring a PPM process. In this article, we suggest that a better approach is to focus on the defectives produced and compare them in as many ways as possible with good units. The key is to find out how the defectives are different from good units in terms of some other characteristic (or combination of characteristics). Identifying such differences can provide a continuous measurement that can be monitored effectively using $\bar{X}$ and R control charts.

This article is organized in the following manner. In the next section previously suggested control chart based approaches are criticized. We demonstrate the weaknesses of using a p-chart to monitor the defective rate directly or Shewhart and sequential charts for the time between defectives. Next a possible remedy is discussed. The approach is based on using a retrospective analysis to identify explanatory variates whose values are related to the defective rate. An example from the automotive industry is given.

2 Critique of Previously Proposed Approaches

One approach to monitor the process performance of a PPM process is to use control charts on either the number of defectives or the time between defectives. Control charting has a long and successful history. The idea is that by quickly determining when the process performance deteriorates, the cause of the deterioration can be identified and eliminated. Control charts work well when the process output can be measured on some continuous scale or when the defective rate is not close to the extremes 0 or 1.

To implement any control chart, we must first observe the process in an in-control state for long enough to allow us to estimate the in-control process performance fairly accurately. However, with PPM processes there is not enough information in the occasional defective item either to set up a control chart or to use it effectively.

2.1 Control Chart to Monitor the Proportion Defective

In the case of good/defective process output, a p chart is recommended (Montgomery, 1991). Subgroups of size n are taken periodically from the process and the proportion of defectives is recorded. To set up a *p*-chart, the standard rule is to set the control limits at $p_0 \pm 3\sqrt{\left[p_0(1-p_0)\right]/n}$, where p_0,

the in-control proportion defective is estimated as $\bar{p}$, the average defective rate in numerous subgroups from an in-control process.

These control limits, based on a normal approximation, are not totally satisfactory if np_0 is small as is likely the case in a PPM environment. The limits may be improved either by using an *arcsin* transformation or better yet, by using probability limits derived from the binomial distribution (Ryan, 1989).

However, even with probability limits, there are two major difficulties in using and setting up a *p* chart here. First, to obtain any reasonable power to detect changes in the PPM process, large subgroup sizes are needed. Montgomery (1991) recommended subgroup sizes large enough so that the probability of finding at least one defective in the subgroup is at least 95%. For a defective rate of p_0 this requirement translates into a subgroup size larger than $3/p_0$. When p_0 is small this rule of thumb results in enormous sample sizes. For example, assuming the process produces 50 PPM defectives when in control, then the $3/p_0$ rule implies subgroups of a least 60,000 units. A second problem is that to set up the chart, we need to estimate p_0 accurately. Assuming we follow the standard recommendation to collect initially at least 20 subgroups in order to set up the chart, in the above example, we must inspect 1.2 million units (from an in-control process) before we can begin to monitor. Similarly, when the defective rate is 5 PPM, the minimum sample size is 600,000 units and 18 million units are needed just to set up the control chart! In most applications, we suspect that these numbers are so large as to make the procedure inoperable.

2.2 Control Chart to Monitor the Time Between Defectives

A clever idea to alleviate the discreteness inherent in monitoring the number of defectives is to monitor the time (or number of good units) between observed defective units. Denote the time between observed defectives as X. In this way, we change the problem from one with discrete measurements to one with a continuous scale. Note that we also avoid the difficult problem of subgroup definition. This approach makes most sense for processes subject to 100% inspection.

Using the time between defectives as a test statistic, we may employ either a Shewhart type chart, or some sequential procedure, such as an EWMA chart or CUSUM chart (Nelson, 1994). This approach was first suggested by Montgomery (1991), and further explored by Nelson (1994), and McCool and Joyner-Motley (1998). Nelson (1994) suggests an individual chart of $X^{.2777}$ (X to the power 0.2777) to monitor the time between defectives. McCool et al. (1998) consider a number of different possible test statistics and control charts. In particular, they suggest that an exponentially weighted moving average chart (EWMA) of $X^{.2777}$ or $log(X)$ would be appropriate.

Unfortunately, there remain inherent difficulties with this approach in the PPM environment. First, it is expensive to perform the 100% inspection of units required to determine the time between defectives. If 100% inspection is

not used the second difficulty is exacerbated. Second, a good estimate of in-control mean time between defectives is needed to set appropriate control limits. For PPM processes, the time between defectives is long, and thus the amount of time (or number of units) required to gather enough information to allow reasonably precise estimation of the mean time between defectives may be too long to be practical. For example, Nelson (1994) suggests that two dozen values of the time between defectives, while the process is in control, are required to reasonably estimate the mean time between failures. When the process produces 5 PPM defective, the expected number of parts between defectives is 200,000. Thus to get 24 values of X requires about 4.8 million units.

Even if we were able to estimate the in-control mean time between defectives accurately the control charts are still not very effective. For example, we may consider some typical results from McCool et al. (1998) for the 5 PPM process. They give the average run length (ARL) for a control chart based on the time between defectives. When the defective rate has increased to 500 PPM the ARL is given as 168.85. However, there are still, on average, 2000 units between failures, therefore on average the chart will pick up a change to 500 PPM defective only after 377,700 units have been inspected. Looking for increases in quality is even worse because the average time between defectives increases. Say the defective rate is reduced to .5 PPM then the ARL is 2.07 and this corresponds to, on average, over 4 million parts!

This problem is not avoided by using a sequential procedure such as an EWMA chart. The performance of an EWMA chart in detecting small shifts will be somewhat superior to the Shewhart chart, but the initial implementation of the EWMA control chart still requires an initial estimate for the in-control defective rate which is not available without massive production volume.

3 A Possible Remedy

The major difficulty in monitoring a PPM process using either the number of defectives in a subgroup or the time between defectives is the small amount of information per unit inspected that such data provides. One solution to this problem is to find an explanatory variate or combination of explanatory variates that are (strongly) related to the defective rate. We assume that the identified explanatory variate is not the underlying continuous measurement that defines defectives and non-defectives, but rather some other product or process characteristic. Naturally, the identified explanatory variate will not be a perfect predictor of whether a unit is defective or not. However, changes in the explanatory variate should be related to the defective rate.

If such an explanatory variate X is found, we can monitor the process using this continuous variate rather than based on the defective rate or time between defectives. This approach avoids the discreteness difficulty in the original problem. The problem is how to identify the important explanatory variable(s) when there are so few defective units. The key is to focus on the defectives that do occur. The defectives are compared to good units on as many process and

product characteristics, denoted $X_1, \ldots, X_k$, as possible. This approach was promoted by Dorian Shainin (see Bhote and Bhote, 2000) and is the same as the idea underlying case/control studies that are widely used to identify risk factors for rare diseases in human populations (Schlesselman, 1982).

To identify important explanatory variates, we model the relationship between the defective rate *p* and the explanatory variates using a logistic regression model, i.e.

$$\begin{aligned} \text{logit}(p) &= \log\left(p/(1-p)\right) \\ &= g(X) = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k \end{aligned} \quad (1)$$

Note that, in this context, often the best model would include non-linear terms, such as $(X_i - t_i)^2$, where t_i is the target value for characteristic X_i. With quadratic terms, any deviation of X_i from its target value will increase the defective rate. Of course interaction terms are also possible. The parameters in the above model $\beta_1, \ldots, \beta_k$ can be estimated using a sample of defective units and a sample of good units using standard approaches (Hosmer et al. 1989). The assumption is that each of these samples is representative of the two types. Based on such data, we can estimate a function that differs from $g(X)$ only in the intercept term. That is, we can estimate

$$\hat{\alpha}_0 + \hat{\beta}_1 X_1 + \ldots + \hat{\beta}_k X_k \quad (2)$$

The intercept term β_0 in $g(X)$ is not estimable without knowledge of the sampling fraction of good and bad units. In fitting this logistic model, the goal is to find explanatory variates whose corresponding model parameters (β_i) are not zero.

3.1 Required Sample Sizes

An important question relates to the number of good (controls) and defective (cases) units that are needed or desirable for the analysis. Clearly when more defectives (and non-defectives) are used in the analysis significant explanatory variates can be identified more easily. However, the number of units required will depend on the magnitude of the effect of the particular explanatory variate.

We explore this issue in more detail by considering a single explanatory variate and assuming the distribution of the explanatory variate is Normal in both the cases and controls. In most work on sample size requirements in case control studies in the medical literature it is assumed that the explanatory variate is binary (Schlesselman, 1982). However, Lubin et al. (1988) consider a continuous explanatory variate. Assuming a single explanatory variate the

sample size required for a logistic regression model to identify an explanatory variate as significant is equivalent to a comparison of the mean levels of the explanatory variate for good and defective units. We let μ_i and σ_i be the mean and standard deviation of the explanatory variate for good units (i = 0) and for defectives (i = 1). We find the sample size needed for a size α-level test with power $1-\beta$.

Following Lubin et al. (1988), the required number of defective units (cases), assuming *k* not defective units (controls) for each case is

$$m = \frac{k+1}{k}\frac{\left[Z_\alpha\sigma_x + Z_\beta\left\{\left(k\sigma_1^2+\sigma_0^2\right)/\left(k+1\right)\right\}^{1/2}\right]^2}{\left(\mu_1-\mu_0\right)^2} \qquad (3)$$

where $\sigma_x^2 = \left(\sigma_1^2+k\sigma_0^2\right)/\left(k+1\right)+\left(\mu_1-\mu_0\right)^2 k/\left(k+1\right)^2$, and Z_α and Z_β satisfy the equations $\Pr\left(Z > Z_\alpha\right)=\alpha$, $\Pr\left(Z > Z_\beta\right)=\beta$ where Z is a standardized Normal random variable.

To explore equation (3) we set α = 0.05 and β = 0.1, and without loss of generality consider the case where for controls the distribution of the explanatory variate is standardized to have mean zero and standard deviation equal to one, i.e. μ_0 = 0 and σ_0 = 1. This leads to three unknowns, the mean and standard deviation for defectives and the number of controls per case we choose. Only the number of controls per case can be chosen by the investigator, so these results should be consider simply guidelines. First, we consider the situation where we have a single control for each case. Figure 1 shows contours of constant *m* as a function of μ_1 and σ_1. We see that, as expected, the number of cases required decreases as the mean of the explanatory variate for cases increases.

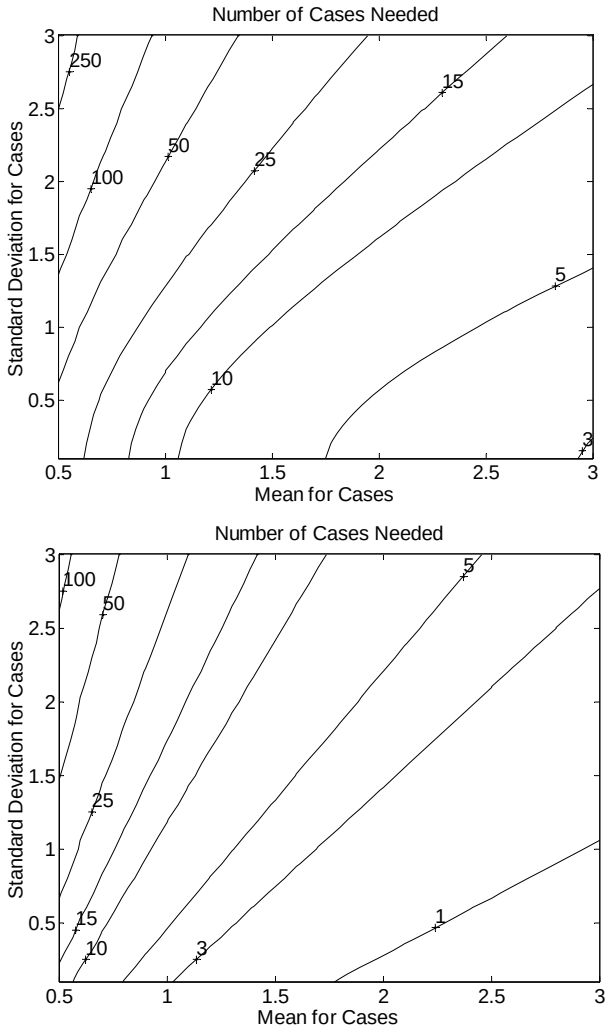

Fig. 1. Contours of the number of defective units needed
one control for each case on top, large number of controls for each case on the bottom

To give a better idea of the effectiveness of increasing the number of controls for each case Figure 2 shows the ratio of the required sample sizes from the two plots in Figure 1, i.e. Figure 2 plots m_1/m_∞, where m_k give the required number of cases when there are *k* not defective units (controls) for each case. It is interesting that the improvement possible through the use of multiple controls is much greater when the explanatory variate is continuous than when the explanatory is binary. Following Schlesselman (1982) using a large number

of controls when the explanatory variate is binary can decrease the number of cases needed by at most a factor of 2. In the industrial PPM context increasing the number of controls (non defective units) for each case (defective units) is typically inexpensive.

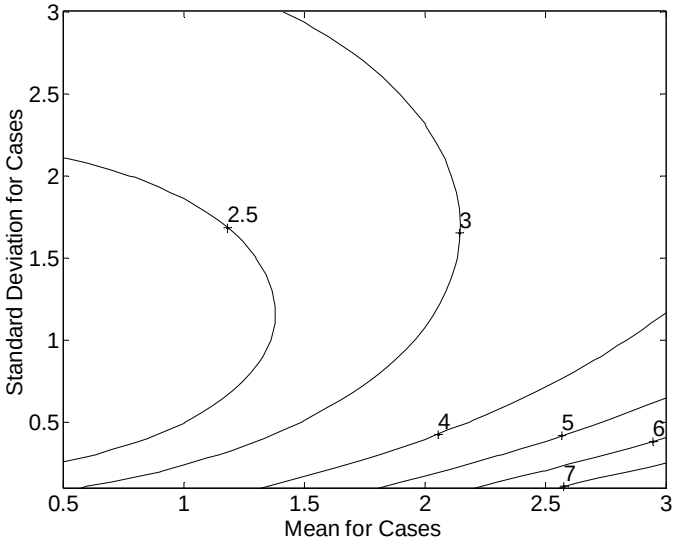

Fig. 2. Contours of the ratio of the required sample size for one control vs. many controls

4 Example

Exhaust valve seats are force fitted by insertion into the head of an engine. If the valve seat is not installed correctly, it can lead to a catastrophic engine failure. The quality of the fit is judged by visual inspection using feeler gauges. Given the high volume (four seats per head, two heads per engine, 1500 engines per day), 100% inspection is very costly and likely to be ineffective, especially since the expected defective rate is low, less than 50 PPM.

A sample of 25 defective seat insertions was collected over time. Pareto analysis showed that there was no evidence that the poorly fitted seats depended on location in the head. Since no head had more than one defective seat, the remaining three seats on the head were used as controls.

Eleven measurable, potentially important explanatory variates were identified. These are X_1, X_2, X_3, X_4, X_5, measurements of force, work and distance taken during the automated insertion process, X_6, X_7, X_8, X_9, dimensional and physical characteristics of the valve seat and X_{10} and X_{11} dimensional characteristics of the pocket in the head into which the seat is inserted.

The explanatory variates X_6 through X_{11} were measured on seats after insertion and there was suspicion that their values may have been distorted by the insertion process. Nevertheless these variates were included in the analysis. The plan was to include those characteristics identified as being important in the monitoring procedure. However, in that case, they would be measured before the insertion process.

A logistic regression model was fit to the 100 observations. Three explanatory variates, X_2, X_4 and X_9 were identified as important. Since X_2 and X_4 were measured on every insertion, an automated CUSUM chart based on a linear combination of X_2 and X_4 was constructed using the software available in the insertion process. The characteristic of the valve seat X_9 was monitored separately using an average and range chart based on subgroups of 5 parts collected with a regular frequency.

5 Discussion

In applying the proposed remedy it is important that the identified explanatory variates used in the linear combination (2) are truly related to the defect rate. If they are not we will be monitoring a linear combination that may well vary in ways unrelated to the defective rate. In this case, the introduction of spurious explanatory variates will "muddy the waters."

A typical reaction in industry to very rare defects is to do a complete "postmortem," including a detailed look for the probably causes, on each individual defective observed. While such an approach can yield improvements it is flawed unless we link the results from all the "postmortems" in an attempt to find some commonality. This idea of looking for commonalities across many defects is one of the principles underlying the proposed approach.

Finally, it should be noted that a designed experiment is not likely to be a viable alternative approach to process improvement in the context described in this article. With a binary response (defective / not defective) either very large sample sizes would be needed, or the factor levels would have to be set so extremely that the conclusions from the experiment would likely not be relevant to the current operating conditions of the process.

6 Summary

Monitoring of the defective rate of a process that produces defectives measured in parts per million (PPM) is often desirable. The use of control charts to monitor the defective rate of such processes based on the number of defectives or the time between defectives is shown to be infeasible. The sample sizes required to set up the control charts are much too large in most practical situations. As an alternative, we suggest a focus on the few defectives that are produced. In particular, we suggest a case/control type comparison of

defectives and non-defectives based on as many of their other attributes as possible. If we can find some explanatory variable or combination of variables that is associated with defectives (or non defectives) we may be able to determine a continuous variate to monitor.

Acknowledgments

This research was supported, in part, by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

References

Bhote, K.R. and Bhote, A.K. (2000), *World Class Quality*, 2nd edition, AMACOM, American Management Association, New York.

Geyer, P.L., Steiner, S.H. and Wesolowsky, G.O. (1996) "Optimal SPRT and CUSUM Procedures with Compressed Limit Gauges," *IIE Transactions*, 28, pp. 489-496.

Hosmer, D.W. Lemeshow, S. (1989), *Applied Logistic Regression*, Wiley and Sons, New York.

Lubin JH, Gail MH, Ershow AG (1988), "Sample Size and Power for Case-Control Studies When Exposures are Continuous," *Statistics in Medicine*, 7, 363-376.

McCool, J. I., and Joyner-Motley, T. (1998), "Control Charts Applicable When the Fraction Nonconforming is Small," *Journal of Quality Technology*, 30, 240-247.

Montgomery, D.C. (1991), *Introduction to Statistical Quality Control*, Second Edition, John Wiley and Sons, New York.

Nelson, L. S., (1994), "A Control Chart for Parts-Per-Million Nonconforming Items," *Journal of Quality Technology*, 26, 239-240.

Ryan, T.P. (1989), *Statistical Methods for Quality Improvement*, John Wiley and Sons, New York.

Schlesselman, J.J. (1982), *Case Control Studies*, Oxford University Press, Oxford.