Monitoring and diagnosing a two-stage production process with attribute characteristics

A. Aghaie¹, Y. Samimi² and S. Asadzadeh³

Multistage process monitoring has recently attracted notable attention in that the statistical relationships between quality variables are taken into account. Here, we dealt with the problem of monitoring and diagnosing a two-stage production process with attribute characteristics in which the outgoing quality variable is impacted by the incoming quality variable from the first process stage. Based on a sampling procedure which inspects each n produced items, these attribute characteristics are assumed to follow binomial and Poisson distributions. Several monitoring techniques including a new method based on the generalized Poisson distribution are presented and the comparison is made to evaluate the effectiveness of these procedures. Moreover, some fault diagnosis methods are fully explored in order to alleviate the identification of the process stage responsible for the out-of-control conditions. The results of the simulation based studies reveal that a combined approach consisting of a proportion defective control chart and an adjusted control chart is quite efficacious in addressing the problems with regard to both the detection power and the fault diagnosis.

Keywords: Multistage processes, Fault diagnosis, Dependent attribute quality characteristics, Generalized Poisson distribution.

1. Introduction

In multivariate statistical process control (SPC), it is required to use process knowledge about the inter-relationships among process parameters in order to establish an appropriate monitoring method. There are different mechanisms leading to association between process quality indicators. Sometimes the process variables are related, but it is possible for any of them to experience a shift without affecting the others. In multistage processes consisting of serial value-added manufacturing operations, a parameter shift in any process variable may affect some or all of the measures in the downstream stages and none of the measures preceding it. While monitoring a multi-stage manufacturing process, there arise some important questions. The first issue involves prompt detection of any out-of-control situation. To propose an appropriate answer to this question requires that the dependency structure across process quality characteristics grouped into successive process stages be determined. Furthermore, as the second issue, when we face an out-of-control situation, the stages and correspondent subset of process variables contributing to the out-of-control condition should be discovered.

Cause-selecting [14] and regression adjusted control charts [1] have been proposed for monitoring and diagnosing multistage processes with normally distributed characteristics. Wade and Woodall use cause-selecting control chart for monitoring two-step processes in which the incoming variable belonging to the first stage can be charted as usual, but the outgoing indicator in the second stage is monitored after adjusting for the incoming quality [14]. Hawkins [1] shows that a regression of each variable, of the upstream stages, on driving variables (downstream variables)

¹Associate Professor, Department of Industrial Engineering, K.N. Toosi University of Technology, Tehran, Iran. aaghaie@kntu.ac.ir
²PhD student, Department of Industrial Engineering, K.N. Toosi University of Technology, Tehran, Iran. y_samimi@dena.kntu.ac.ir
³Ph student, Department of Industrial Engineering, K.N. Toosi University of Technology, Tehran, Iran. sh_asadzadeh@dena.kntu.ac.ir
is able to provide a suitable diagnostic means for monitoring cascade processes. Then, he plots regression residuals on separate individual control charts [2].

The main objective of this paper involves monitoring a two-stage process with attribute quality characteristics. Besides, appropriate methods for the interpretation of an out-of-control signal are discussed. Unlike multivariate normally distributed processes, little work has been carried out yet for multistage processes with attribute quality characteristics. As a matter of fact, monitoring multivariate attribute processes has recently been proposed as an attractive research area of SPC. Apparently, Patel wrote the first paper about SPC for the observations from a multivariate binomial or Poisson distributions [4]. By assuming approximate normality for reasonably large sample of observations, he suggested a $T^2$ control chart. Lu et al. [5] gave a multivariate attribute control chart (MACC) which was a straightforward extension of the univariate np-chart [4]. Niaki and Abbasi [7] proposed two different methods for monitoring multivariate Poisson and binomial process measures [7, 8]. The first method was based on the Bootstrap approach for computing simultaneous confidence intervals. And in the second paper, the power transformation was proposed to make the skewness of the marginal distributions as close to zero as possible. Hence, it would be possible to use the $T^2$ control chart for the transformed variables. It should be noted here that although we do not employ some recent advances in monitoring and diagnosing multi-attribute processes like Niaki and Abbasi [7-10] which were nicely been reviewed by Topalidou and Psarakis [12], we strongly encourage interested readers to examine their performance for monitoring multi-stage attribute processes.

Recently, numerous papers have used regression adjustment approach based on generalized linear models (GLM) for process variables that follow discrete distributions. Skinner et al. [13] developed a procedure for monitoring multiple discrete counts [3]. This procedure uses the likelihood ratio statistic for Poisson counts when input variables are measurable. Jearkpaporn et al. [2] extended this approach for monitoring a tree-stage process whose characteristics are a mixture of normally and non-normally distributed variables [2].

The problem considered in this paper includes a two-stage model with corresponding Bernoulli and Poisson distributed characteristics. The applications of such a two-stage model, which will be formulated in the next section, can be found in many different industries like the following instances:

- In cotton-spinning process, the skein strength depends on the fiber strength. The latter may be gauged as a binary random variable while the number of tears along a given length of skein may quantify the former.
- In the semiconductor manufacturing process, there are three steps: an oxide deposition process, followed first by a spin on glass (SOG) coat, and then by an oxide etch process. Two quality characteristics are the impurity or particle counts and the average of the oxide thickness, where the former may depend in some degrees on the latter.
- For manufacturing glass containers, and also for producing car windshield, the number of surface blemishes like blebs may depend on the properties of the incoming molten glass.

The remainder of this paper is organized as follows. The next section deals with the problem formulation. Section 3 proposes some methods for monitoring methods process quality characteristics. The issue of process fault diagnosis, which involves identifying the sources of process out-of-control conditions, is illustrated in Section 4. Final remarks and conclusion are provided in Section 5.

2. Problem statement

Consider a two-stage process along with their distinct quality characteristics. In the first stage, each inspected item is categorized as conforming or non-conforming on the basis of coincidence
with the given requirements and then in the second stage the number of defects is enumerated for each item. Quality inspection for the products having passed the whole process is performed by gathering a sample of $n$ items in an station located after the second stage. Therefore, both the number of defective items as well as the total number of defects are computed, by considering respective requirements of each stage. Assuming that the Poisson distribution can represent conditional variability of the second characteristic given the value of the first random variable, we continue our discussion by presenting a probabilistic model for the process.

To examine the likely dependence of product characteristics across process stages, consider Figure 1, which provides a suitable description of the propagation of the product quality through consecutive stages. Figure 1 depicts the empirical distribution of the number of defects on each inspected unit with regard to its status in the first stage. Accordingly, a hypothesis test is run to verify the initial guess about the dependence of product quality measures across two process stages.

![Figure 1. Relationship between quality characteristics of the consecutive stages of the process](image)

Minitab’s analysis output to test the null hypothesis of the equality of defect rates is shown in Table 1. This result, which is based on the pooled estimate of the defect rate, clearly infers the significant dependence of the number of defects in the second stage on the status of the product in the first stage.

Based on this illustration, to formulate the process, a homogenous Bernoulli distribution will be used to represent the variability of the first characteristic with the known parameter $p_0$, while the process is supposedly operating in-control. Hence, assuming that consecutive products are independently distributed, we can model the number of defective products in the sample of $n$ items as a binomial random variable with parameters $n$ and $p_0$. Additionally, two Poisson random variables with different rates $\lambda_{10}$ and $\lambda_{20}$ are needed to represent the variability of the number of defects respectively for non-defective and defective items generated in the first stage. Figure 2 shows a formulation of the process.

Here, we aim at developing an appropriate method for monitoring process quality of the process depicted above. Our first objective involves detecting shifts in model parameters $p_0$, $\lambda_{10}$, and $\lambda_{20}$. In case of an out-of-control situation, the second objective involves identifying the peculiar stage of the process contributing to the change. Sections 3 and 4 discuss respectively these two issues.
Table 1. Testing the hypothesis of the equality of defect rates in the second stage for non-defective and defective items

**Test and CI for Two-Sample Poisson Rates: Count of defects, defective/non-defective**

<table>
<thead>
<tr>
<th>Count of defects</th>
<th>Total Occurrences</th>
<th>Rate of occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>420</td>
<td>0.5009</td>
</tr>
<tr>
<td>1</td>
<td>326</td>
<td>1.06286</td>
</tr>
</tbody>
</table>

\[
\text{Difference} = \text{rate(0)} - \text{rate(1)}
\]

Estimate for difference: -1.86277

95% CI for difference: (-1.56176, -1.16377)

Test for difference = 0 (vs not = 0): Z = -10.83 P-Value = 0.000

\[
\text{Exact Test: P-Value } = 0.000
\]

produced in the first stage.

**Figure 2.** Probability model of a two-stage process.

3. Process monitoring

This section delineates several methods for monitoring the two-stage process introduced in Section 2. The \( np \) and \( c \) control charts are two widely used SPC charts for monitoring attribute characteristics that follow respectively the binomial and Poisson probability distributions [6]. The \( np \) control chart monitors the number of defective items in a sample and the \( c \) control chart is employed to draw the number of defects observed on a given inspection unit. Regardless of the dependence between successive stages of the process, it seems that \( np \) and \( c \) control charts are reasonable options for process monitoring. Hence, with a sample of \( n \) items after the second stage, the number of non-conforming items of the first stage can be drawn on the \( np \) control chart with limits defined below:

\[
CL_{np} = np_0,
\]

\[
UCL_{np}, LCL_{np} = np_0 \pm z_{\alpha/2} \sqrt{np_0(1-p_0)},
\]

where \( n \) denotes the sample size and \( z_{\alpha/2} \) indicates the \( 100(1-\alpha/2) \) percentage point of the standard normal distribution. It should be noted that \( \alpha \), probability of error type I, is usually
Monitoring and diagnosing a two-stage production

determined so that a desired in-control average run length is obtained. Average run length indicates the average number of points plotted on the control chart until an out-of-control signal.

In the second stage, the average number of defects on each inspection unit can be determined by equation (2)
\[ \lambda_{y_0} = (1 - p_0)\lambda_{10} + p_0\lambda_{20}. \]  
(2)

Accordingly, a \( c \) control chart used for monitoring the total number of defects for a sample of size \( n \) is defined as follows.
\[ CL_c = n\lambda_{y_0}, \]
\[ UCL_c, LCL_c = n\lambda_{y_0} \pm z_{\alpha/2}\sqrt{n\lambda_{y_0}}, \]  
(3)

where \( \alpha_2 \) denotes the probability of error type I.

In order to draw control charts in equations (1) and (3), \( \alpha_1 \) and \( \alpha_2 \) must be specified. Following the Bonferroni inequality, \( \alpha_1 \) and \( \alpha_2 \) can be chosen so that the total probability of error type I, as the result of parallel use of two control charts, does not exceed some desired value, \( \alpha \). If we construct each control chart based on equal \( \alpha/2 \) probabilities of error type I, we would meet the condition. For instance, to obtain the total probability of 0.0027, a common value for Shewhart control charts, \( \alpha_1 \) and \( \alpha_2 \) are obtained as 0.00135 which if subtracted from 1 yields 0.9987 as the in-control probability for each control chart.

It should be noticed that the actual value of error type I can be computed via the joint distribution of the number of defective items, \( D \), and the number of defects, \( C \), in a sample of \( n \) products. The following equation shows the details of the computation:
\[ \alpha = 1 - Pr\left[\left(LCL_{np} < D < UCL_{np}\right) \cap \left(LCL_c < C < UCL_c\right)\right] \]
\[ = 1 - \sum_{x=LCL_{np}}^{[UCL_{np}]} \left(\frac{n}{x}\right)p_0^x(1-p_0)^{n-x}\left[\sum_{y=LCL_c}^{[UCL_c]} e^{-[(n-x)\lambda_{10} + x\lambda_{20}]}\left\{(n-x)\lambda_{10} + x\lambda_{20}\right\}^y/y!\right] \]  
(4)

where \( \langle LCL \rangle \) means the smallest integer greater than or equal to \( LCL \) and denotes the largest integer less than or equal to \( UCL \). Based on different values for the first parameter, \( p_0 \), the fraction of defective items produced in the first stage, Table 2 contains the overall probability of error type I. It should be added that hypothetical values of the defect rates \( \lambda_{10} \) and \( \lambda_{20} \), and also the sample size, \( n \), have been considered 0.5, 2 and 100 respectively in our calculations.

For the \( c \) control chart, the second column includes \( (1 - \alpha_2) \) that is close to its nominal value; i.e., 0.9987, when \( p_0 \) approaches zero or one. The fourth column of Table 2 indicates that the total amount of \( \alpha \) for the combination of control charts may even reach 0.0107. This means that while the process is operating in-control, an erroneous out-of-control signal may appear after each 93.46 periods on the average, though the ARL is expected to be 370.37. As a matter of fact, because the number of defects in the second stage is over-dispersed, the rate of false alarms increases considerably. Therefore, it is anticipated to have frequent out-of-control points on the \( c \) control chart, whereas there are no assignable causes. This issue will be elaborated in the next subsection.

3.1. Modified \( c \) control chart: \( c' \) control chart

The problem of excessive false alarms has been caused by the incorrect computation of control limits. Although obvious, it is worth to note that the variability of the number of defects is
influenced by the variation of the number of defective items in the first stage. This extra variation was overlooked when we calculated the control limits of $c$ chart in equation (3). Equation (5) indicates the proper value of variance for the number of defects in each item observed in the second stage:

$$
\sigma_{\gamma_0}^2 = (1 - p_0)\lambda_{10}^2 + p_0\lambda_{20}^2 - [ (1 - p_0)\lambda_{10} + p_0\lambda_{20} ]^2 + (1 - p_0)\lambda_{10} + p_0\lambda_{20}.
$$

(5)

<table>
<thead>
<tr>
<th>$p_0$</th>
<th>(1 - $\alpha$) for the $c$ control chart</th>
<th>(1 - $\alpha$) for the combination of $np$ and $c$ control charts</th>
<th>Total $\alpha$ for the combination of $np$ and $c$ control charts</th>
<th>Nominal variance</th>
<th>Real variance</th>
<th>Increase in the variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.9984</td>
<td>-</td>
<td>-</td>
<td>50.00</td>
<td>50.00</td>
<td>0.00%</td>
</tr>
<tr>
<td>0.0</td>
<td>0.9959</td>
<td>0.9920</td>
<td>0.0080</td>
<td>57.50</td>
<td>68.19</td>
<td>18.59%</td>
</tr>
<tr>
<td>0.1</td>
<td>0.9939</td>
<td>0.9922</td>
<td>0.0078</td>
<td>65.00</td>
<td>85.25</td>
<td>31.15%</td>
</tr>
<tr>
<td>0.1</td>
<td>0.9924</td>
<td>0.9912</td>
<td>0.0088</td>
<td>72.50</td>
<td>101.19</td>
<td>39.57%</td>
</tr>
<tr>
<td>0.2</td>
<td>0.9916</td>
<td>0.9902</td>
<td>0.0098</td>
<td>80.00</td>
<td>116.00</td>
<td>45.00%</td>
</tr>
<tr>
<td>0.2</td>
<td>0.9914</td>
<td>0.9900</td>
<td>0.0100</td>
<td>87.50</td>
<td>129.69</td>
<td>48.21%</td>
</tr>
<tr>
<td>0.3</td>
<td>0.9916</td>
<td>0.9904</td>
<td>0.0096</td>
<td>95.00</td>
<td>142.25</td>
<td>49.74%</td>
</tr>
<tr>
<td>0.3</td>
<td>0.9901</td>
<td>0.9893</td>
<td>0.0107</td>
<td>102.50</td>
<td>153.69</td>
<td>49.94%</td>
</tr>
<tr>
<td>0.4</td>
<td>0.9910</td>
<td>0.9899</td>
<td>0.0101</td>
<td>110.00</td>
<td>164.00</td>
<td>49.09%</td>
</tr>
<tr>
<td>0.4</td>
<td>0.9921</td>
<td>0.9907</td>
<td>0.0093</td>
<td>117.50</td>
<td>173.19</td>
<td>47.39%</td>
</tr>
<tr>
<td>0.5</td>
<td>0.9916</td>
<td>0.9909</td>
<td>0.0091</td>
<td>125.00</td>
<td>181.25</td>
<td>45.00%</td>
</tr>
<tr>
<td>0.5</td>
<td>0.9929</td>
<td>0.9915</td>
<td>0.0085</td>
<td>132.50</td>
<td>188.19</td>
<td>42.03%</td>
</tr>
<tr>
<td>0.6</td>
<td>0.9929</td>
<td>0.9916</td>
<td>0.0084</td>
<td>140.00</td>
<td>194.00</td>
<td>38.57%</td>
</tr>
<tr>
<td>0.6</td>
<td>0.9943</td>
<td>0.9933</td>
<td>0.0067</td>
<td>147.50</td>
<td>198.69</td>
<td>34.70%</td>
</tr>
<tr>
<td>0.7</td>
<td>0.9945</td>
<td>0.9931</td>
<td>0.0069</td>
<td>155.00</td>
<td>202.25</td>
<td>30.48%</td>
</tr>
<tr>
<td>0.7</td>
<td>0.9958</td>
<td>0.9941</td>
<td>0.0059</td>
<td>162.50</td>
<td>204.69</td>
<td>25.96%</td>
</tr>
<tr>
<td>0.8</td>
<td>0.9961</td>
<td>0.9944</td>
<td>0.0056</td>
<td>170.00</td>
<td>206.00</td>
<td>21.18%</td>
</tr>
<tr>
<td>0.8</td>
<td>0.9972</td>
<td>0.9958</td>
<td>0.0042</td>
<td>177.50</td>
<td>206.19</td>
<td>16.16%</td>
</tr>
<tr>
<td>0.9</td>
<td>0.9976</td>
<td>0.9956</td>
<td>0.0044</td>
<td>185.00</td>
<td>205.25</td>
<td>10.95%</td>
</tr>
<tr>
<td>0.9</td>
<td>0.9979</td>
<td>0.9937</td>
<td>0.0063</td>
<td>192.50</td>
<td>203.19</td>
<td>5.55%</td>
</tr>
<tr>
<td>1</td>
<td>0.9987</td>
<td>-</td>
<td>-</td>
<td>200.00</td>
<td>200.00</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

The fifth and the sixth columns of Table 2 contain the nominal value of the variance of the number of defects for all $n$ items in a sample, in conjunction with the real variance that is obtained by multiplying equation (5) by $n$. The percentage of increase in the variance with respect to the nominal variance is shown in the last column. It is obvious that the largest amount of the
probability of error type I corresponds to the largest difference between nominal and real variance. According to equation (5), the variance of the number of defects approaches its nominal value, $\lambda_{Y_0}$, only when $p_0$ equals zero or one. Consequently, the control limits of the $c'$ chart are modified as follows:

$$CL_{c'} = n\lambda_{Y_0},$$
$$UCL_{c'}, LCL_{c'} = n\lambda_{Y_0} \pm z_{a/2}\sigma_{Y_0}. \quad (6)$$

Henceforth, the $c'$ chart will be used instead of the initial $c$ chart as the supplement to the $np$ chart defined in equation (1).

### 3.2. New np control chart for defective items: np$_d$ control chart

Here, we reconsider the definition of a non-defective item. Let a non-defective item mean an item which possesses neither of the flaws of the process. Therefore, the probability of observing a defective item can be computed as follows.

$$p_{d_0} = p_0 + (1 - p_0)(1 - e^{-\lambda_{0}}). \quad (7)$$

Thus, a new $np$ chart for monitoring total number of defective items can be a reasonable alternative for control charting process quality. This chart will be called the $np_d$ control chart. Control limits of $np_d$ chart are given by equation (8).

$$CL_{np'} = np_{d0}$$
$$UCL_{np'}, LCL_{np'} = np_{d0} \pm z_{a/2}\sqrt{np_{d0}(1 - p_{d0})} \quad (8)$$

The $np_d$ chart will be used alone. Therefore, the total number of defective items based on the revised definition is plotted on the chart. Since here we need just one control chart, the probability of error type I, $\alpha$, can be simply determined using the standard normal distribution.

### 3.3. Employing the generalized Poisson distribution: GPD control chart

The total number of defects observed in a sample of $n$ items does not follow the Poisson distribution, though its conditional distribution is assumed to do so. Since the number of defects on each unit depends on the unit’s status from the preceding stage, the variance of the number of defects differs significantly from its mean unless $p_0$ is zero or one. However, the generalized Poisson distribution (GPD) may be regarded a suitable alternative to model variability of the total number of defects in a sample. This distribution has been employed by He et al. (2006) to represent the distribution of the number of defects for a high-quality process, where the over-dispersion of data had to be taken into account [13]. The probability mass function of the GPD is shown by equation (9) below:

$$f_z(\theta, \mu) = \left[\theta(\theta + z\mu)^{\mu-1} e^{-\theta - z\mu}\right] z!; \quad z = 0,1,2,..., \mu \geq 0. \quad (9)$$

Positive values of $\mu$ accounts for over-dispersed data. The mean and variance of GPD are given as $E(Z) = \theta/(1 - \mu)$ and $Var(Z) = \theta/(1 - \mu)^3$, respectively.

Recall that the mean and variance of the number of defects calculated in equations (2) and (5). Therefore, parameters of the GPD may be specified by solving the following system of equations,

$$\{E(Z) = n\lambda_{Y_0}, \quad Var(Z) = n\sigma_{Y_0}^2, \quad (10)$$

which results in $\mu = 1 - \sqrt[\lambda_{Y_0}/\sigma_{Y_0}^2}$ and $\theta = n\lambda_{Y_0}\sqrt[\lambda_{Y_0}/\sigma_{Y_0}^2]$. 


For example, suppose that while the process is in-control, process parameters are $p_0 = 0.2$, $\lambda_{10} = 0.5$, and $\lambda_{20} = 2$. If the sample size is 100, then the corresponding GPD parameters are obtained as $\mu_0 = 0.17$ and $\theta_0 = 66.56$. Additionally, to check the adequacy of the GPD, a test procedure for evaluation of the goodness-of-fit based on the chi-square statistic can be performed. The $p$-value of 0.62 assures us that the GPD could appropriately represent the variability of the number of defects in a sample of $n$ products.

Using cumulative distribution function of GPD, control limits may be obtained easily. Hence, the probability limits for this control chart that will be called GPD chart, are found by solving the following equations.

$$F(LCL_{GPD}; \theta_0, \mu_0) = \sum_{z=0}^{LCL_{GPD}} f_Z(\theta_0, \mu_0) \leq \alpha/2,$$

$$F(UCL_{GPD}; \theta_0, \mu_0) = \sum_{z=0}^{UCL_{GPD}} f_Z(\theta_0, \mu_0) \geq (1 - \alpha/2).$$

(11)

Therefore, as long as the number of defects for each sample is plotted inside in-control region, we cannot reject the assumption that the process is in-control.

3.4. Evaluation of the monitoring methods

To evaluate the sensitivity of the proposed methods, the operating characteristic (OC) curve of each one is drawn and compared with that of other methods based on a simulation study that follows the model formulation in Section 2. It is worthy of attention that the evaluation made in this section is based on a particular study in which the in-control values of process parameters are $p_0 = 0.2$, $\lambda_{10} = 0.5$, and $\lambda_{20} = 2$. Figures 3-5 show the OC curves of control charts versus the different values of parameters when $\alpha$ is fixed at 0.0027.

Here, exact probabilities for the combination of $np$ and $c'$ control charts, were obtained by the joint distribution of the number of defective items and the number of defects. On the other hand, corresponding values for $np_d$ and GPD control charts were obtained by using the cumulative distribution function of binomial distribution and GPD respectively.

![Figure 3: Operating characteristic (OC) curve of the control charts versus the mean of the Bernoulli random variable (parameter of the first stage) while $p_0 = 0.2$.](image-url)
With regard to changes in the first parameter, Figure 3 shows that the np control chart combined with the c’ control chart outperforms the other two control charts which exhibit quite a similar performance. On the other hand, as shown by Figure 4, the np_d control chart is the most sensitive control chart with respect to the changes in parameter λ_10. To detect decreasing shifts of λ_10, the GPD control chart stands in the second position as compared with the np_d chart. Nevertheless, there is no noticeable difference between the GPD and the np - c’ control charts in detection of increasing shifts in λ_10. Concerning the third parameter, Figure 5 infers a similar interpretation about the performance of the GPD and the np - c’ control charts. The most obvious result of this figure concerns the inability of the np_d control chart to detect shifts of λ_20.

**Figure 4.** Operating characteristic (OC) curve of the control charts versus the rate of the first Poisson random variable (parameter of the second stage) while λ_10 = 0.5.

**Figure 5.** Operating characteristic (OC) curve of the control charts versus the rate of the second Poisson random variable (parameter of the second stage) while λ_20 = 2.
4. Fault diagnosis

Section 3 proposed three approaches to detect shifts in process parameters. Identifying the location of shift that is called process fault diagnosis is another problem which is discussed in this section. As illustrated by equation (2), in the two-stage process considered here, shifts in the parameter of the first stage of the process may transfer to the second stage. Therefore, we will not be able to have appropriate interpretation of the location of the shift for an out-of-control signal even using combined $np - c'$ control charts that employ separate charts for process stages.

For the purpose of process fault diagnosis, this section compares two different approaches. The first approach, which is essentially used for fault diagnosis in multivariate SPC, assumes that process variables follow the multivariate normal distribution. In our study, the sample size is supposed to be large enough to provide an approximate bivariate normal distribution for our discrete binomial and Poisson random variables. This approach is illustrated in the next subsection. The second approach, is described in subsection 4-2, is based on the adjustment of the mean number of defects in the second stage according to the number of defective items in the first stage.

4.1. Decomposition of the chi-square statistic

A brief description of the chi-square control chart is needed to illustrate our first approach for diagnosis. Suppose there are $k$ related quality characteristics which are assumed to follow jointly a $k$-variate normal distribution. Let the vector of sample means of these quality characteristics be denoted by $\bar{x} = [\bar{x}_1, \bar{x}_2, ..., \bar{x}_k]'$. For each sample, the test statistic which is plotted on the chi-square control chart is computed as $\chi^2 = n(\bar{x} - \mu)' \Sigma^{-1} (\bar{x} - \mu)$, where $\mu$ represents process mean and $\Sigma$ denotes the covariance matrix. This test statistic follows a chi-square distribution with $k$ degrees of freedom, and so the upper control limit is specified as $2k \chi^2{\alpha'}$. In case of unknown parameters, $\mu$ and $\Sigma$ can be estimated retrospectively by analysis of preliminary observations.

Consider sample statistics represented by $\bar{x}$ and $S$. Then, the procedure is called the $T^2$ control chart. Exact control limit for $T^2$ statistic has been reported in well-known SQC textbooks. When estimated parameters are obtained through a large number of preliminary samples, the upper control limit of the chi-square control chart is considered by an approximate value. The $T^2$ approach was firstly used by Patel [3] as an approximate method for monitoring quality characteristics that follow multivariate binomial or multivariate Poisson distributions. Here the chi-square control chart is used for monitoring simultaneously two characteristics, the first of which has a binomial distribution and the second is a Poisson distributed variable.

By now, various approaches have been proposed for the interpretation of an out-of-control signal in $T^2$ control chart. A suitable approach which is adopted here is based on the decomposition of the monitoring statistic. If we suppose that $T^2$ statistic is available for a sample of observations, and $T^2_{(j)}$ is the value of the statistic for all variables except the $j$th variable, then Runger et al. [11] proposed that $d_j = T^2 - T^2_{(j)}$ can measure the contribution of the $j$th variable to the overall $T^2$ statistic. When one faces an out-of-control signal, it is recommended to assess the values of $d_j$ to determine most contributing variables. An approximate cut-off value for $d_j$ can be $\chi^2_{\alpha',1}$, as reported by Montgomery [6]. In what follows, we set $\alpha'$ at 0.05. It should be added that there are some other interesting and rather more computationally intensive approaches to decompose $T^2$ statistic.
According to the process model formulation in Section 2, the covariance of the two process variables is determined by equation (12) blow:

$$\sigma_{XY} = p_0\lambda_{20} - p_0\left[ p_0\lambda_{20} + (1 - p_0)\lambda_{10} \right]$$

(12)

Hence, the covariance matrix of the process is obtained as follows:

$$\Sigma = \begin{bmatrix} p_0(1 - p_0) & \sigma_{XY} \\ \sigma_{XY} & \sigma_{Y0}^2 \end{bmatrix},$$

(13)

where $\sigma_{Y0}^2$ and $\sigma_{XY}$ are given by equations (5) and (12), respectively. Consequently, for a sample of size $n$ the approximate joint distribution of the number of defective items and the number of defects in a sample of size $n$ may be presented as given by equation (14) below:

$$D, C \sim N(\left[ np_0, n(p_0\lambda_{20} + (1 - p_0)\lambda_{10}) \right], \Sigma),$$

(14)

where $D$ and $C$ denote two process random variables. The chi-square control statistic then can be used to monitor the mean vector of the process. In case of any out-of-control signal, the root-cause analysis is conducted based on the decomposition of the control statistic as elaborated above.

4.2. Adjustment of the limits of the control chart for the number of defects

The second approach for fault diagnosis is based on the adjustment of the mean number of defects in the second stage of the process according to the number of defective items produced in the first stage. Similar to the first monitoring approach introduced in subsection 2-1, two supplementary control charts are employed here. However, to prevent the second control chart from being affected by the shifts in the first parameter, it is required to modify the control limits of the $c$ chart in the second stage as shown by equation (15) blow:

$$CL_c(t) = \left[ D(t)\lambda_{20} + (n - D(t))\lambda_{10} \right],$$

$$UCL_c(t), LCL_c(t) = CL_c(t) \pm z_{\alpha/2} \sqrt{CL_c(t)}.$$

(15)

To avoid confusion, this control will be called the adjusted $c$ control chart and will be denoted as the $c_{adj}$ chart. As shown by equation (15), the $c_{adj}$ control chart that accompanies the $np$ chart, possesses a moving center line which is adjusted based on the number of defective items in sample $t, D(t)$.

Figure 6. Control charts for a two-stage process whose first parameter has changed.
To clarify the difference between this new approach and the previous approach of combined np - c’ control charts, consider Figure 6 that pertains to a process whose first parameter has shifted at the 10th period from $p_0 = 0.2$ to $p_1 = 0.3$. The np control chart has produced immediately an out-of-control signal. An increasing trend has also emerged on the c’ control chart though there is no change in the parameters of the second stage. However, the adjusted c control chart has adapted itself to the change in $p_0$, and consequently there is no out-of-control signal appeared on this control chart.

The results of the first row must be interpreted carefully. Concerning the shifts in the first parameter, $p_0$, the first approach that is based on the decomposition of $T^2$ statistic can successfully identify the location of change. As an illustration, Figure 8 shows that when $p_0$ changes from in-control value of, 0.2 to 0.4, the distribution of the number of defects is apparently affected (note the dotted line that represents zero). However, the volume of information provided by $d_2 = T^2 - T_{(2)}^2$ is masked with the magnification of the squared values of $T_{(2)}^2$ which leads to small values for $d_2$. This issue is more clearly explained by Figure 9. Remarkably, the distribution of the adjusted number of defects retains its original shape against shift in $p$, so that the standardized distribution resembles the presumed standard normal distribution, Figure 8.

Most interestingly, Figure 7-b of which is also confirmed by the last part in the third row, reveals that the adjusted c control chart can successfully detect the location of change. Conversely, the ability of $T^2$ decomposition method decreases as the shift size increases. In effect, for the large values of $\lambda_1$ and $\lambda_2$, both components of $T^2$ statistic resulted from the decomposition of the overall statistic, will trigger a warning signal, though the change occurred only in the second stage as detected properly by the adjusted c control chart.

![Figure 7a. np - c_adj method along with $T^2$ method to detect shifts in $p$ ($p_0 = 0.2$) in the first stage](image-url)
**Figure 7b.** \( np \cdot c_{adj} \) method along with \( T^2 \) method to detect shifts in \( \lambda_1 \) (\( \lambda_{10} = 0.5 \)) in the second stage.

**Figure 7c.** \( np \cdot c_{adj} \) method along with \( T^2 \) method to detect shifts in \( \lambda_2 \) (\( \lambda_{20} = 2 \)) in the third stage.

**Figure 7.** Identification of the process stage suffering from shift using proposed approaches for fault diagnosis.
Figure 8. Empirical frequency distribution of the standardized values with $p_1 = 0.4$. of process statistics when the parameter of the first stage shifts from $p_0 = 0.2$ to $p_1 = 0.4$.

Figure 9. Frequency distribution of the statistics resulted from the $T^2$ decomposition when the parameter of the first stage shifts from $p_0 = 0.2$ to $p_1 = 0.4$.

5. Conclusion
The main application area of our findings is multistage manufacturing processes whose quality characteristics are countable attributes. Considering the growing interest in monitoring multistage
processes, it seems quite important to study the multiple attribute manufacturing processes where characteristics are related hierarchically through cascade process stages. The most fundamental achievements in this context involves not only monitoring methods, which provide prompt detection of process shifts, but also appropriate diagnosing techniques, which aid in process root-cause analysis. We investigated a two-stage production process with corresponding attribute characteristics for each process stage. In the first stage, the conformance of each product to some requirements is inspected while in the second stage the number of defects on each inspected item is determined.

Three different approaches for monitoring the process was proposed. Accurate understanding of the probability distributions of process characteristics provided the basis for developing monitoring schemes. A modification of the Shewhart \( c \) control chart to account for the over dispersion of the number of defects in the second stage, along with the new kind of Shewhart \( np \) control chart and also another approach based on the generalized Poisson distribution (GPD), constitute the contributions for process monitoring.

Diagnosing the process through interpretation of an out-of-control signal as an important issue in multistage manufacturing processes was also examined. Although the fault diagnosis methods of multivariate normal processes were applicable by approximation, it indicated that a rather simple approach based on the adjustment of the control limits of the chart in the second stage could be strongly beneficial in detecting the location of the assignable causes. A simulation-based comparison study clarified the superiority of the adjustment method over the approximate method based on the decomposition of the \( T^2 \) statistic. In conclusion, the monitoring techniques and diagnosing approaches proposed and studied here could be quite effective in root-cause analysis of the multistage manufacturing processes with attribute quality characteristics. And finally, as it was mentioned before, we emphasize that using a multi-attribute control charting method along with its fault classification procedure can be an interesting idea to be used for monitoring multi-stage attribute processes. Comparison of the performances of the proposed methodology with that of the existing multi-attribute monitoring procedures in terms of both out-of-control average run length and the correct classification percentages can be concerns stage for the future work.

Acknowledgements
The authors express their thanks to the referees for their valuable comments and helpful suggestions.

References


