

Research Article

A Nonparametric Scheme for Monitoring a Process Output with a Block Effect

Saad T. Bakir

College of Business Administration, Alabama State University, 915 South Jackson Street, Montgomery, AL 36104, USA

Correspondence should be addressed to Saad T. Bakir, bakir00@yahoo.com

Received 28 May 2012; Revised 3 August 2012; Accepted 17 August 2012

Academic Editor: Shuen L. Jeng

Copyright © 2012 Saad T. Bakir. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposes a distribution-free (or nonparametric) control scheme to monitor a process output that contains two special causes of variation called “block or batch” effects and “treatment or position” effects. The scheme properties (control limits, false alarm rate, and in-control average run length) stay the same under any assumed continuous probability distribution. For moderate sample sizes, these properties can be computed exactly from available tables without the need to estimate the mean or variance of the process. The proposed monitoring scheme requires ranking the observations within blocks and using the method of analysis of means by ranks. The paper includes an illustrative example concerning the grinding process of silicon wafers used in integrated circuits production.

1. Introduction

In statistical process control, there are instances in which the process output contains block effects component in addition to the treatment effects component that is to be controlled. In manufacturing integrated circuits on silicon wafers, Roes and Does ([1, Table 1]) reported data bearing two effects: the batch (or block) effect and the position (or treatment) effect of the wafer under the grinder. To take account of both effects, they proposed a Shewhart-type control chart that is based on a two-way analysis of variance (ANOVA) model for controlling the process treatment mean and for controlling certain linear contrasts of the wafer positions. In manufacturing paper or plastic films, Palm and DeAmico [2] reported that the raw manufacturing material was formed into a continuous sheet (web) which would then be wound up into rolls at the end of the production line. Because of nonuniformities in the product along the cross-direction perpendicular to the direction of the travel of the web, samples were taken from each roll at different cross-directional positions (e.g., the middle, the front or operator side, and the back or motor side.) Palm and DeAmico [2] modeled the data as a two-way ANOVA setup where rolls served as blocks and cross-directional positions served as treatments. To account for the block effects, they suggested performing

periodic cross-directional studies that are based on the analysis of main effects (ANOME) when monitoring the silicon oxide on computer chips, Yashchin [3] reported that from each lot, a sample of wafers is selected and then several measurements are made on each of the selected wafers. Based on a nested random effects model, he developed a cumulative sum (CUSUM) control chart for monitoring the process variance components. Earlier, Woodall and Thomas [4] considered monitoring variance components in a nested random effects model situation.

All the above monitoring schemes are distribution based (or parametric) that assume a normal distribution for the process output. The normality assumption, however, may not be valid for some manufacturing data; citations are given in Bakir [5], Chakraborti and Graham [6], and Chakraborti et al. [7].

The purpose of this paper is to develop a distribution-free (or nonparametric) control scheme for monitoring a process output that is under the influence of two special causes of variation called “block or batch” effects and “treatment or position” effects. While accounting for the block component, the proposed scheme monitors and detects differences in the treatment components of the process output. Our approach involves ranking the raw measurements of the process output within blocks and then using the method of analysis of means

by ranks (ANOMR) as developed in Bakir [8]. The proposed scheme has the following merits.

- (a) The scheme takes account of the variability in the data that is due to block effects.
- (b) The scheme is distribution-free because its properties (control limits, false alarm rate, and in-control average run length) stay the same under any continuous probability distribution.
- (c) For moderate sample sizes, the scheme properties can be computed exactly without the need to estimate the mean or the standard deviation of the process. For large sample sizes, large sample approximations are available.

2. A Nonparametric Monitoring Scheme

The purpose of the monitoring scheme is to monitor a process output that contains two special causes of variation called “block or batch” effects and “treatment or position” effects. Examples of such output were cited in the Introduction.

2.1. Sampling Plan. At each time instance we take a sample of n blocks (e.g., rolls of plastic film or batches), then we make k measurements (e.g., at fixed positions on the roll or the batch) on the process output within each of the n blocks. Thus, a total of nk observations x_{tij} ($i = 1, 2, \dots, n; j = 1, 2, \dots, k$) is gathered at each sampling time instance t , $t = 1, 2, \dots$

2.2. Model and Assumptions. For such x_{tij} data, an appropriate model is the following two-way ANOVA model:

$$X_{tij} = \mu + \beta_{i(t)} + \tau_{tj} + \varepsilon_{tij} \quad \text{for } t = 1, 2, \dots; \quad (1)$$

$$i = 1, 2, \dots, n; \quad j = 1, 2, \dots, k.$$

In model (1), at each sampling instance t , $t = 1, 2, \dots$, we have that X_{tij} is the measurement on the j th treatment in the i th block, μ is an unknown constant representing the process overall mean, $\beta_{i(t)}$ is an unknown constant representing the effect of the i th block nested within time, τ_{tj} is an unknown constant representing the fixed effect of the j th treatment, and ε_{tij} is a random variable having a continuous probability distribution with zero median; it represents common cause variation in the (tij) th observation. The ε'_{tij} s ($t = 1, 2, \dots$, $i = 1, 2, \dots, n$, and $j = 1, 2, \dots, k$) are assumed mutually independent.

2.3. Quality Characteristic to Be Monitored. The purpose of the monitoring scheme is to detect differences among the treatment components while accounting for the block effects. This translates to the process being in control if the hypothesis, H_0 (called the in-control hypothesis), holds true for all t , $t = 1, 2, \dots$

$$H_0 : \tau_{t1} = \tau_{t2} = \dots = \tau_{tk}. \quad (2)$$

The out-of-control hypothesis is

$$H_1 : \text{the } \tau\text{'s are not all equal.} \quad (3)$$

2.4. Charting Statistic. At each sampling instance t , $t = 1, 2, \dots$ and for $i = 1, 2, \dots, n$, let R_{tij} denote the rank of X_{tij} within the i th block $\{X_{ti1}, X_{ti2}, \dots, X_{tik}\}$. Formally,

$$R_{tij} = 1 + \sum_{s=1}^{j-1} I(X_{tis} < X_{tij}), \quad (4)$$

where $I(\cdot)$ is the indicator function.

Calculate the treatment rank totals

$$R_{t \cdot j} = \sum_{i=1}^n R_{tij} \quad \text{for } t = 1, 2, \dots, \quad j = 1, 2, \dots, k. \quad (5)$$

It can be verified that under the in-control hypothesis H_0 in (2), the rank totals, $R_{t \cdot j}$, have mean μ and variance σ^2 given by

$$\mu = \frac{n(k+1)}{2}, \quad \sigma^2 = \frac{n(k+1)(k-1)}{12}. \quad (6)$$

Define the treatment rank deviations

$$D_{t \cdot j} = R_{t \cdot j} - \mu \quad \text{for } t = 1, 2, \dots, \quad j = 1, 2, \dots, k. \quad (7)$$

We will use the $D_{t \cdot j}$'s as charting statistics for a follow-up monitoring scheme of the process. Define the maximum absolute rank deviations

$$D_t = \max_{1 \leq j \leq k} |D_{t \cdot j}| \quad \text{for } t = 1, 2, \dots \quad (8)$$

We will use D_t as a charting statistic for an initial monitoring scheme.

2.5. The In-Control Probability Distribution of the Charting Statistic, D_t . At each sampling instance t , $t = 1, 2, \dots$ and under the in-control hypothesis H_0 in (2), the $k!$ possible rank configurations $(R_{ti1}, R_{ti2}, \dots, R_{tik})$ are equally likely within each of the n blocks. Since the blocks are independent, there are $(k!)^{n-1}$ equally likely rank configurations at each sampling time instance. Using these facts, Bakir ([8, Table A.1]) tabulates the exact in-control distribution of D_t for the cases when $k = 3$ and $n = 3(1)9$; $k = 4$ and $n = 3(1)6$; and when $k = 5$ and $n = 3, 4$. Bakir's table gives the values for α and $d_{\alpha; k, n}$ satisfying

$$\Pr(D_t \geq d_{\alpha; k, n} | H_0) = \alpha, \quad (9)$$

D_t is a nonparametric test statistic because its distribution (under H_0) is the same for any parent continuous process distribution. The large sample in-control distribution of D_t is discussed in Bakir [8] and its upper percentile points are given in Bakir ([9, Table IV]).

2.6. Signaling Rule and Control Limits

2.6.1. Initial Monitoring. First, we start an initial monitoring scheme by plotting the maximum absolute rank deviation, D_t , against time $t = 1, 2, \dots$. For a desired prechosen probability of false signal, α , the scheme signals at the first sampling time instance t , $t = 1, 2, \dots$ for which

$$D_t \geq d_{\alpha; k, n}. \quad (10)$$

TABLE 1: Values of UCL, FAR, and ARL₀.

Number of blocks, n	Number of treatments, k								
	$k = 3$			$k = 4$			$k = 5$		
	UCL	FAR	ARL ₀	UCL	FAR	ARL ₀	UCL	FAR	ARL ₀
$n = 3$	3	.1944	5.3	4.5	.1181	8.5	5	.2800	3.6
							6	.0075	133.3
$n = 4$	4	.0694	14.4	5	.1418	7.1	6	.2120	4.70
				6	.0307	32.6	7	.0769	13.0
							8	.0159	62.9
$n = 5$	4	.1242	8.1	5.5	.1452	6.9			
	5	.0239	41.8	6.5	.0451	22.2			
				7.5	.0078	128.2			
$n = 6$	4	.1840	5.4	6	.1443	6.9			
	5	.0521	9.2	7	.0518	19.3			
	6	.0081	23.5	8	.0135	74.1			
				9	.0019	526.3			
$n = 7$	4	.0854	11.7						
	5	.0207	37.0						
	6	.0027	370.4						
$n = 8$	5	.1197	8.4						
	6	.0375	26.7						
	7	.0080	125.0						
	8	.0009							
$n = 9$	5	.1540	6.5						
	6	.0570	17.5						
	7	.0158	63.3						
	8	.0030	333.3						
	9	.0003	3333.3						

The initial monitoring scheme has one control limit given by

$$UCL = d_{\alpha;k,n}, \tag{11}$$

where $d_{\alpha;k,n}$ is defined in (9).

A signal by this initial monitoring only indicates an out-of-control condition; it does not, however, pinpoints the treatment components that triggered the signal. To achieve this objective, we need to perform a follow-up monitoring.

2.6.2. Follow-Up Monitoring. In the follow-up monitoring scheme we plot the individual rank deviations $D_{t,j}$ ($j = 1, 2, \dots, k$), in (7), across time. A signal is given at the first time t , $t = 1, 2, \dots$ for which

$$D_{t,j} \geq d_{\alpha;k,n} \quad \text{or} \quad D_{t,j} \leq -d_{\alpha;k,n}, \tag{12}$$

for at least one $j = 1, 2, \dots, k$.

The upper and lower control limits for this follow-up monitoring scheme, respectively, are

$$UCL = d_{\alpha;k,n}, \quad LCL = -d_{\alpha;k,n}. \tag{13}$$

The signaling rules in (10) and (12) are equivalent in the sense that one signals if and only if the other signals too. For large values of n , the value of $d_{\alpha;k,n}$ may be replaced in all

the formulae that appear in this section by $\sigma\omega(\alpha;k)$, where $\omega(\alpha;k)$ are the upper percentile points of the large sample distribution of D_t given in Bakir ([9, Table IV.])

3. The Average Run Length

Let T be a discrete random variable denoting the time (number of sampling time instances) required until the monitoring scheme's first signal; that is, when $D_t \geq d_{\alpha;k,n}$. T is called the run length of the scheme and its expected value, $E(T)$, is called the average run length (ARL). Since the random variables $D_t, t = 1, 2, \dots$ are independent, T follows a geometric distribution with parameter (probability of a signal), p :

$$p = \Pr(D_t \geq d_{\alpha;k,n}). \tag{14}$$

By properties of the geometric distribution, we get

$$ARL = E(T) = \frac{1}{p}. \tag{15}$$

The monitoring scheme is nonparametric because its in-control (under H_0) run length distribution is geometric with the same parameter p for any parent continuous process distribution. When the process is in control, let ARL_0 denote

ARL and let FAR denote the false alarm rate (probability of a false signal). Then

$$FAR = \Pr(D_t \geq d_{\alpha;k,n} | H_0) = \alpha, \quad ARL_0 = \frac{1}{\alpha}. \quad (16)$$

It is clear that ARL_0 and FAR do not depend on the functional form of the probability distribution of the process. Table 1 gives the exact values of FAR, ARL_0 , and the corresponding UCL for some values of n and k . Table 1 was generated by using results in Bakir ([8, Table A.1]). For large values of n , one may use the large sample distribution of D_t to compute ARL_0 .

4. Illustrative Example

We apply the proposed monitoring scheme to data relating to the grinding process of silicon wafers used in integrated circuits production. The wafers are positioned in a circular arrangement under a grindstone to achieve a certain desired thickness. Thickness measurements (μm) are taken on wafers seated on five fixed positions (labeled P1, P2, P3, P4, and P5) under the grinder. The P1 and P2 positions fall on the outer circle, P3 and P4 in the middle, and P5 in the inner circle. Thickness data of 30 successive batches of wafers are given in Roes and Does ([1, Table 1]). To apply our scheme, we break the data into ten ($t = 10$) time (sampling) instances with three ($n = 3$) batches (blocks) at each time. Measurements are made at the five ($k = 5$) positions (treatments) within each block.

Formulas in (6) give the following values for the mean and variance of the position (treatment) rank totals, $R_{t,j}$:

$$\mu = \frac{3(5+1)}{2} = 9.0, \quad \sigma^2 = \frac{3(5+1)(5-1)}{12} = 6.0. \quad (17)$$

Table 2 displays the thickness data, assignments of the appropriate ranks, and further computations of the rank totals for the five positions (P1, P2, ..., P5). Using Table 1 with $n = 3$, $k = 5$, and FAR of $\alpha = 0.0075$ (equivalent to $ARL_0 = 133$), we read the upper control limit: $UCL = 6$. Figure 1 displays the resulting chart for the initial monitoring scheme and Figures 2 and 3 show the resulting charts for the follow-up monitoring scheme.

Figure 1 shows that, except for the fifth sampling time instance, the successive values of the charting statistic, D_t , fall right on the upper control limit. Thus, the initial monitoring scheme triggers an out-of-control signal indicating significant variation among the wafers' positions. However, Figure 1 does not pinpoint which position or positions caused the signal. We proceed now to perform the follow-up monitoring scheme as shown in Figures 2 and 3.

Figure 2 is a multiline chart showing the individual position profiles across time. This follow-up scheme signals at each time except for the fifth when all positions become within the control limits. Moreover, Figure 2 pinpoints the positions that caused the signal; namely, P1 and P4. Positions 2, 3, and 5 are within the control limits at all sampling times.

The follow-up scheme in Figure 3 is a single-line chart showing the individual position profiles across time. Figure 3

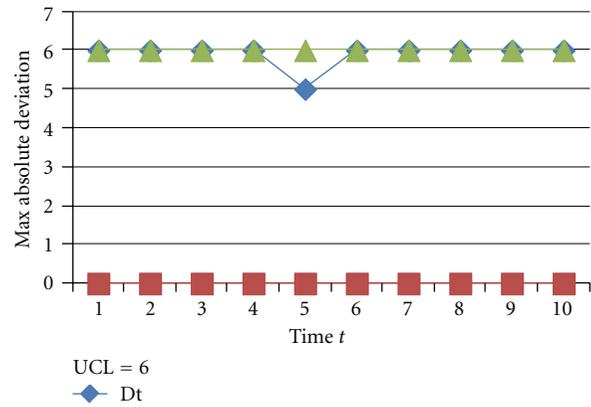


FIGURE 1: Initial Monitoring Chart.

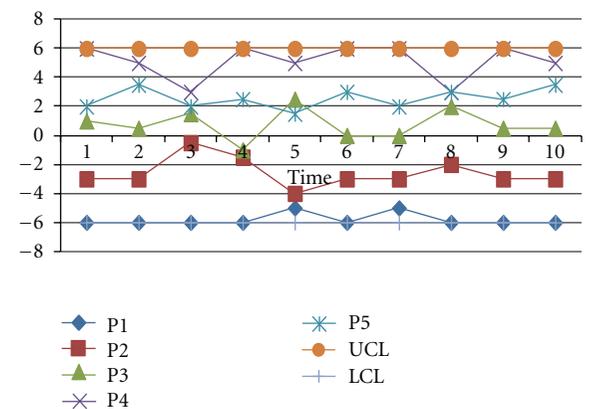


FIGURE 2: Follow-up Multi-line Monitoring Chart.

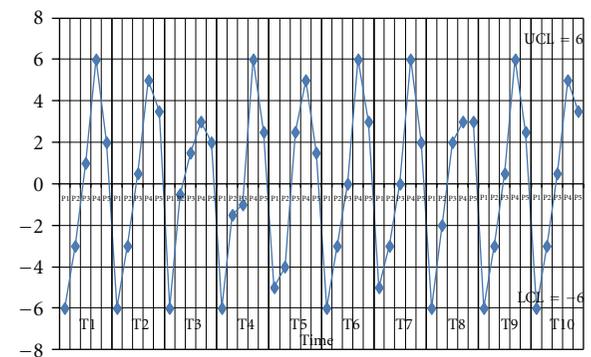


FIGURE 3: Follow-up Single-line Monitoring Chart.

leads to the same conclusions as in Figure 2. Charting the follow-up monitoring scheme by either Figure 2 or Figure 3 is a matter of choice by the practitioner. It is interesting to note that Figure 3 resembles the chart of the famous Red Bead Experiment of Deming ([10], Figure 38 page 347). In the Red Bead Experiment, the numbers of defectives produced by each one of six workers are plotted over four days. The workers correspond to the wafer positions in our current example.

TABLE 2: Thickness (μm) of wafers. P indicates position of wafer.

Time t	Block i	P1	Rank	P2	Rank	P3	Rank	P4	Rank	P5	Rank
1	1	240	1	243	2	250	4	253	5	248	3
1	2	238	1	242	2	245	3	251	5	247	4
1	3	239	1	242	2	246	3	250	5	248	4
	$R1,j$		3		6		10		15		11
	$D1,j$		-6		-3		1		6		2
2	1	235	1	237	2	246	3.5	249	5	246	3.5
2	2	240	1	241	2	246	3	247	4	249	5
2	3	240	1	243	2	244	3	248	5	245	4
	$R2,j$		3		6		9.5		14		12.5
	$D2,j$		-6		-3		0.5		5		3.5
3	1	240	1	243	2	244	3	249	5	246	4
3	2	245	1	250	4.5	250	4.5	247	2	248	3
3	3	238	1	240	2	245	3	248	5	246	4
	$R3,j$		3		8.5		10.5		12		11
	$D3,j$		-6		-0.5		1.5		3		2
4	1	240	1	242	2	246	3	249	5	248	4
4	2	240	1	243	2	246	3	250	5	248	4
4	3	241	1	245	3.5	243	2	247	5	245	3.5
	$R4,j$		3		7.5		8		15		11.5
	$D4,j$		-6		-1.5		-1		6		2.5
5	1	247	2	245	1	255	5	250	4	249	3
5	2	237	1	239	2	243	3	247	5	246	4
5	3	242	1	244	2	245	3.5	248	5	245	3.5
	$R5,j$		4		5		11.5		14		10.5
	$D5,j$		-5		-4		2.5		5		1.5
6	1	237	1	239	2	242	3	247	5	245	4
6	2	242	1	244	2	246	3	251	5	248	4
6	3	243	1	245	2	247	3	252	5	249	4
	$R6,j$		3		6		9		15		12
	$D6,j$		-6		-3		0		6		3
7	1	243	1	245	2	248	3	251	5	250	4
7	2	244	1	246	3	246	3	250	5	246	3
7	3	241	2	239	1	244	3	250	5	246	4
	$R7,j$		4		6		9		15		11
	$D7,j$		-5		-3		0		6		2
8	1	242	1	245	2	248	3	251	5	249	4
8	2	242	1	245	3	248	5	243	2	246	4
8	3	241	1	244	2	245	3	249	5	247	4
	$R8,j$		3		7		11		12		12
	$D8,j$		-6		-2		2		3		3
9	1	236	1	239	2	241	3	246	5	242	4
9	2	243	1	246	2	247	3.5	252	5	247	3.5
9	3	241	1	243	2	245	3	248	5	246	4
	$R9,j$		3		6		9.5		15		11.5
	$D9,j$		-6		-3		0.5		6		2.5
10	1	239	1	240	2	242	3	243	4	244	5
10	2	239	1	240	2	250	3.5	252	5	250	3.5
10	3	241	1	243	2	249	3	255	5	253	4
	$R10,j$		3		6		9.5		14		12.5
	$D10,j$		-6		-3		0.5		5		3.5

5. Concluding Remarks

In this paper, we propose a distribution-free (or nonparametric) scheme for monitoring treatment components while accounting for block effects in a process output. The procedure is based on a nonparametric version of the analysis of means method. Being distribution-free, the FAR and the ARL_0 , of the procedure stay the same under any assumed continuous distribution for the process. For small sample sizes and small number of treatments, the control limits, FAR and ARL_0 , can be computed exactly from available tables without the need to estimate the mean or variance of the process. The paper includes an example concerning the grinding process of silicon wafers used in integrated circuits production.

References

- [1] K. C. B. Roes and R. J. M. M. Does, "Shewhart-type charts in nonstandard situations," *Technometrics*, vol. 37, pp. 15–24, 1995.
- [2] A. C. Palm and R. L. DeAmico, "Shewhart-type charts in nonstandard situations: discussion," *Technometrics*, vol. 37, pp. 26–29, 1995.
- [3] E. Yashchin, "Monitoring variance components," *Technometrics*, vol. 36, no. 4, pp. 379–393, 1994.
- [4] W. H. Woodall and E. V. Thomas, "Statistical process control with several components of common cause variability," unpublished paper presented at the Joint Statistical Meetings, Atlanta, Ga, USA, 1991.
- [5] S. T. Bakir, "A distribution-free Shewhart quality control chart based on signed-ranks," *Quality Engineering*, vol. 16, no. 4, pp. 613–623, 2004.
- [6] S. Chakraborti and M. A. Graham, "Nonparametric control charts," in *Encyclopedia of Quality and Reliability*, John Wiley, New York, NY, USA, 2007.
- [7] S. Chakraborti, S. W. Human, and M. A. Graham, "Nonparametric (distribution-free) quality control charts," in *Handbook of Methods and Applications of Statistics: Engineering, Quality Control, and Physical Sciences*, N. Balakrishnan, Ed., pp. 298–329, John Wiley & Sons, New York, NY, USA, 2011.
- [8] S. T. Bakir, "Analysis of means using ranks for the randomized complete block design," *Communications in Statistics—Simulation*, vol. 23, no. 2, pp. 547–568, 1994.
- [9] S. T. Bakir, "Analysis of means using ranks," *Communications in Statistics—Simulation*, vol. 18, pp. 757–776, 1989.
- [10] W. E. Deming, *Out of the Crisis*, MIT, Cambridge, Mass, USA, 1986.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

