NIASRA

NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

National Institute for Applied Statistics Research Australia

University of Wollongong

Working Paper

13-15

Statistical Process Control for Autocorrelated Data on Grid

Anoop Chaturvedi, Ashutosh Kumar Dubey and Chandra Gulati

National Institute for Applied Statistics Research Australia, University of Wollongong, Wollongong NSW 2522. Phone +61 2 4221 5435, Fax +61 2 4221 4845. Email: anica@uow.edu.au

Statistical Process Control for Autocorrelated Data on Grid

Anoop Chaturvedi
Department of Statistics
University of Allahabad, India
anoopchaturv@gmail.com

Ashutosh Kumar Dubey
Department of Statistics
University of Allahabad, India
dubeyakk@gmail.com

Chandra Gulati
National Institute for Applies Statistics Research, Australia
School of Mathematics and Applied Statistics
University of Wollongong, Australia
cmg@uow.edu.au

July 23, 2015

Abstract

Assumptions of normality and independence are key assumptions made in the development and use of control charts. However for much industrial data, the assumption of independence is hard to justify. Control charts for the sample mean and sample standard deviation are obtained, when the observations are taken over a grid on a two-dimensional surface. It is assumed that the observations follow an autoregressive process of order one in both the directions. The results have been applied to road pavement data on road corse thickness observed on a two dimensional grid.
Key words: Autocorrelated data; $\bar{x}$ chart; s chart; AR(1) process.

1 Introduction

The statistical control charts developed by Shewart (1931) have wide applications in industry for monitoring and improving the manufacturing process. A common assumption underlying development and use of these control charts is that observations are uncorrelated. However, it has been observed in several manufacturing processes that this assumption is not satisfied and successive observations show serial correlation. As pointed out by Porter and Caulcutt (1992) and Roes and Does (1995), the standard control charts are applicable under restrictive assumptions, which are often not satisfied in various applications. Thus the control chart has to be modified to the nature of production process at hand. Roes and Does (1995) proposed method for control charting the data generated by mixed model and applied the proposed procedure to data on thickness of grinding wafers in integrated cicuits production of Philips plant in Netherlands. Montgomery and Mastrangelo (1991) observed that the presence of autocorrelation among observations leads to false alarms and misleading conclusions about the control state of the process. Even small and positive autocorrelation drastically reduces the in-control average run length of the chart. Zhang (1998a) considered EWMAST chart for weakly stationary data. Process capability indexes for the autocorrelated processes have been studied by Zhang (1998b). Control charts for autocorrelated data based on generalized likelihood ratio are discussed in Capizzi and Masarotto (2008).

Two alternative approaches to handle the problem of autocorrelation in data are commonly employed: (i) modifying Shewart control chart limits by taking into account the autocorrelation, (ii) applying control charts to estimated residuals of the fitted time series model. Wieringa (1999) investigated the performance of these two approaches and observed that residual charts perform better for negative autocorrelation whereas modified control limit charts perform better for positive autocorrelation. He also proposed a modified control chart which outperforms both of these charts. Sparks (2000) extended the one sided CUSUM procedure for controlling autocorrelation in AR(1) processes and compared it with two versions of Shewart charts. Jiang (2001) computed average run length of stationary ARMA charts and, on the basis of simulations, observed that ARMA charts are comparable to the optimal EWMA chart for monitoring IID processes. Zhang (2002) discussed different procedures for dealing with process autocorrelation when using process control charts and process capability indexes.

While monitoring a road construction project of NSW, Australia, Ollis (1997) and Griffiths *et al.* (2003) observed data on base corse thickness of road on a two-dimensional grid, in which there is autocorrelation in both of the orthogonal directions. For such autocorrelated observations, they developed a method to construct control charts for the mean on a two-dimensional grid. Sparks and Ollis (2001) applied universal kriging model for the spatial data and developed control limits to monitor the construction process of new road pavements. However, they did not consider the control chart for standard deviation (s.d.), which plays an important role in process variability control. The control chart for s.d. also needs to be designed accommodating autocorrelation structure of the process on grid.

Using the framework of Griffiths *et al.* (2003) the present paper develops both mean and s.d. charts for autocorrelated observations on a two-dimensional grid. For illustration purpose, considering road pavement data on road corse thickness observed over two dimensional grid (see Griffiths *et al.*, 2003) the mean and standard deviation control charts are plotted using modifying control limits based on AR(1) process in both the directions. Though we have applied the theoretical results to road construction example, the results are applicable to more processes such as monitoring rolling steel/metal to a target thickness, image analysis of spatial quality of a product manufactured over time.

2 Autocorrelated Observations on 2-D Grid

Suppose the sampling of points has been carried out on a two dimensional (2-D) grid of $n = u.v$ points $(t, k), (t = 1, 2, \ldots, u; k = r, 2r, \ldots, vr)$, taken in mutually orthogonal directions. Here u is the number of rows and v is the number of columns on the grid. The distance between two neighboring cells in the direction of rows is taken as unity whereas the distance between two neighboring cells in the direction of columns is taken as r. Further, $x_{t,k}$ represents the observation on $(t, k)^{th}$ cell $(t = 1, 2, \ldots, u; k = r, 2r, \ldots, vr)$. It has been assumed that the observations follow $AR(1)$ process in both the directions, given by

$$\begin{aligned} x_{t,k} &= \mu + \phi_1 x_{t-1,k} + \varepsilon_{t,k} \\ x_{t,k} &= \mu + \phi_2 x_{t,k-r} + \xi_{t,k} \end{aligned} \tag{2.1}$$

where $\varepsilon_{t,k}$ and $\xi_{t,k}$ are *iid* random variables following normal distribution with mean 0 and variances $\sigma_x^2/(1-\phi^2)$ and $\sigma_x^2/(1-\phi^{2r})$ and σ_x^2 is the variance of $x_{t,k}$'s $(0 < \sigma_x^2 < \infty)$. Further, μ, ϕ_1 and ϕ_2 and are the parameters of the model. Let us assume that

$$\phi_2 = \phi_1^{sr}$$

so that

$$s = \frac{\ln \phi_2}{r \ln \phi_1} \tag{2.2}$$

It is now possible to change the scale in the direction of columns and consider the distance between neighboring cells as sr. Then the problem reduces to the problem with same autocorrelation coefficient in two directions. Hence, without loss of generality, we assume that $\phi_1 = \phi$ and $\phi_2 = \phi^r$. We require the transformation of parameters from (ϕ_1, ϕ_2) to (ϕ, s) to accommodated the spatial autocorrelation among observations while constructing control limits.The model (2.1), in terms of new parameters (ϕ, s) , reduces to

$$\begin{aligned} x_{t,k} &= \mu + \phi x_{t-1,k} + \varepsilon_{t,k} \\ x_{t,k} &= \mu + \phi^r x_{t,k-r} + \xi_{t,k} \end{aligned} \tag{2.3}$$

The distance between the points from where $x_{t,k}$ and $x_{t+h,k+l}$ have been taken, denoted by $d(h, l)$ is given by $d(h, l) = \sqrt{(h^2 + l^2 r^2)}$. Hence, the correlation coefficient between $x_{t,k}$ and $x_{t+h,k+l}$ is $\phi^{d(h,l)}$.

In practice, if ϕ_1 and ϕ_2 are unknown, we may estimate the value of s by replacing ϕ_1 and ϕ_2 by their estimators in the expression (2.2) for s.

3 Control Chart for Mean and Standard Deviation

Let us write $x^{(k)} = (x_{1,k}\ x_{2,k} \ldots x_{u,k})'; k = 1, 2, \ldots, v$ and

$$x = \begin{pmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(v)} \end{pmatrix}$$

For obtaining the variance covariance matrix of the random vector x, we observe that

$$\begin{aligned} V(x^{(k)}) &= \sigma_x^2 \begin{pmatrix} 1 & \phi & \ldots & \phi^{u-1} \\ \phi & 1 & \ldots & \phi^{v-2} \\ \vdots & \vdots & \ddots & \vdots \\ \phi^{u-1} & \phi^{u-2} & \ldots & 1 \end{pmatrix} \\ &= \sigma_x^2 \Sigma^{(0)} \qquad \text{(say).} \end{aligned} \tag{3.1}$$

The covariance matrix between the vectors $x^{(k)}$ and $x^{(k+l)}$ is obtained as

$$cov(x^{(k)}, x^{(k+l)}) = \sigma_x^2 \begin{pmatrix} \phi^{d(0,l)} & \phi^{d(1,l)} & \dots & \phi^{d(u-1,l)} \\ \phi^{d(1,l)} & \phi^{d(0,l)} & \dots & \phi^{d(u-2,l)} \\ \vdots & \vdots & \ddots & \vdots \\ \phi^{d(u-1,l)} & \phi^{d(u-2,l)} & \dots & \phi^{d(0,l)} \end{pmatrix} \tag{3.2}$$
$$= \sigma_x^2 \Sigma^{(l)} \qquad \text{(say)}.$$

Then the covariance matrix of the random vector x is given by

$$V(x) = \sigma_x^2 \begin{pmatrix} \Sigma^{(0)} & \Sigma^{(1)} & \dots & \Sigma^{(v-1)} \\ \Sigma^{(1)} & \Sigma^{(0)} & \dots & \Sigma^{(v-2)} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma^{(v-1)} & \Sigma^{(y-2)} & \dots & \Sigma^{(0)} \end{pmatrix} \tag{3.3}$$
$$= \sigma_x^2 \Omega \qquad \text{(say)}.$$

Now, the mean and variance of n observations $x_{t,,k} (t = 1, 2, \dots u; k = r, 2r, \dots, vr)$ are given by

$$\bar{x}_n = \frac{1}{n} l_n' x$$

$$S_n^2 = \frac{1}{n-1} x' A x$$

where l_n is a $n \times 1$ vector with all elements one and

$$A = I_n - \frac{1}{n} l_n l_n'.$$

We can easily verify that $A l_n = 0$.

3.1 Control Chart for Mean

To obtain control limits for the $\bar{x}$-chart, we observe that

$$\begin{aligned} Var(\bar{x}_n) &= \frac{\sigma_x^2}{n^2} l_n' \Omega l_n \\ &= \frac{\sigma_x^2}{n^2} \left[v l_u' \Sigma^{(0)} l_u + 2 \sum_{j=1}^{v-1} (v-j) l_u' \Sigma^{(j)} l_u \right] \\ &= \frac{\sigma_x^2}{n} \left[1 + \frac{2}{u} \sum_{i=1}^{u-1} (u-i) \phi^i + \frac{2}{v} \sum_{j=1}^{v-1} (v-j) \phi^j + \frac{4}{n} \sum_{i=1}^{u-1} \sum_{j=1}^{v-1} (u-i)(v-j) \phi^{d(i,j)} \right] \end{aligned} \tag{3.4}$$

Hence, the lower and upper 3σ control limits for the $\bar{X}$ chart are given by

$$\begin{aligned} LCL_\phi &= \bar{\bar{x}}_n - 3 \frac{s_n}{\sqrt{n}} \left[1 + \frac{2}{u} \sum_{i=1}^{u-1} (u-i) \phi^i + \frac{2}{v} \sum_{j=1}^{v-1} (v-j) \phi^j + \frac{4}{n} \sum_{i=1}^{u-1} \sum_{j=1}^{v-1} (u-i)(v-j) \phi^{d(i,j)} \right]^{1/2} \\ CL_\phi &= \bar{\bar{x}}_n \\ UCL_\phi &= \bar{\bar{x}}_n + 3 \frac{s_n}{\sqrt{n}} \left[1 + \frac{2}{u} \sum_{i=1}^{u-1} (u-i) \phi^i + \frac{2}{v} \sum_{j=1}^{v-1} (v-j) \phi^j + \frac{4}{n} \sum_{i=1}^{u-1} \sum_{j=1}^{v-1} (u-i)(v-j) \phi^{d(i,j)} \right]^{1/2} \end{aligned} \tag{3.5}$$

In practice, if ϕ is unknown, we may replace it by its estimator.

3.2 Control Chart for Standard Deviation

For obtaining the control limits of control chart for standard deviation, we have

$$
\begin{aligned}
E(S_n^2) &= \frac{1}{n-1}E(x'Ax) \\
&= \frac{1}{n-1}trE(Axx') \\
&= \frac{1}{n-1}tr\left[A\left(\sigma^2\Omega + \mu^2 l_n l_n'\right)\right] \\
&= \frac{\sigma_x^2}{n-1}tr(A\Omega) \\
&= \frac{\sigma_x^2}{n-1}\left(tr\Omega - \frac{1}{n}l_n'\Omega l_n\right) \\
&= \frac{\sigma_x^2}{n-1}\left[n - \frac{v}{n}l_u\Sigma^{(0)}l_u - \frac{2}{n}\sum_{j=1}^{v-1}(v-j)l_u\Sigma^{(j)}l_u\right] \\
&= \frac{\sigma_x^2}{n-1}\left[n - 1 - \frac{2}{u}\sum_{i=1}^{u-1}(u-i)\phi^i - \frac{2}{v}\sum_{j=1}^{v-1}(v-j)\phi^j - \frac{4}{n}\sum_{i=1}^{u-1}\sum_{j=1}^{v-1}(u-i)(v-j)\phi^{d(i,j)}\right]
\end{aligned} \tag{3.6}
$$

Again

$$
\begin{aligned}
E(S_n^4) &= \frac{1}{(n-1)^2}E(x'Ax)^2 \\
&= \frac{\sigma_x^4}{(n-1)^2}\left[(trA\Omega)^2 + 2tr(A\Omega A\Omega)\right] \\
&= \left[E(S_n^2)\right]^2 + Var(S_n^2)
\end{aligned} \tag{3.7}
$$

Here the variance of S_n^2 is given by

$$
\begin{aligned}
Var(S_n^2) &= \frac{2\sigma_x^4}{(n-1)^2}tr(A\Omega A\Omega) \\
&= \frac{2\sigma_x^4}{(n-1)^2}\left[tr\Omega^2 - \frac{2}{n}l_n'\Omega^2 l_n + \frac{1}{n^2}\left(l_n'\Omega l_n\right)^2\right]
\end{aligned} \tag{3.8}
$$

Again, we have

$$
tr\Omega^2 = n + 2v\sum_{i=1}^{u-1}(u-i)\phi^{2i} + 2u\sum_{j=1}^{v-1}(v-j)\phi^{2j} + 4\sum_{i=1}^{u-1}\sum_{j=1}^{v-1}(u-i)(v-j)\phi^{2d(i,j)} \tag{3.9}
$$

Further, if ω_{ij} is the $(i,j)^{th}$ element of Ω , then

$$
l_n'\Omega^2 l_n = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\omega_{ik}\omega_{kj} \tag{3.10}
$$

From (3.4), (3.9) and (3.10), substituting the values of different expressions in (3.8), we obtain

$$Var(S_n^2) = \frac{2\sigma_x^4}{(n-1)^2}\left[n + 2v\sum_{i=1}^{u-1}(u-i)\phi^{2i} + 2u\sum_{j=1}^{v-1}(v-j)\phi^{2j} + 4\sum_{i=1}^{u-1}\sum_{j=1}^{v-1}(u-1)(v-j)\phi^{2d(i,j)}\right.$$
$$- \frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\omega_{ik}\omega_{kj} + \left\{1 + \frac{2}{u}\sum_{u-i}^{u-1}(u-i)\phi^{i} + \frac{2}{v}\sum_{j=1}^{v-1}(v-j)\phi^{j}\right.$$
$$\left.\left. + \frac{4}{n}\sum_{i=1}^{u-1}\sum_{j=1}^{v-1}(u-i)(v-j)\phi^{d(i,j)}\right\}^2\right] \tag{3.11}$$

Utilizing (3.6), (3.7) and (3.11), we can obtain the value of $E(S_n^4)$.
Following Zhang (1998b, p. 565), we have

$$E(S_n) \approx \{E(S_n^2)\}^{1/2} \tag{3.12}$$

$$Var(S_n) \approx \frac{Var(S_n^2)}{4E(S_n^2)} \tag{3.13}$$

Utilizing (3.12) and (3.13), we obtain 3σ control limits for control chart for standard deviation as

$$\begin{aligned} LCL_\phi &= E(S_n) - 3\sqrt{Var(S_n)} \approx \{E(S_n^2)\}^{1/2} - 3\sqrt{\frac{Var(S_n^2)}{4E(S_n^2)}} \\ CL_\phi &= E(S_n) \approx \{E(S_n^2)\}^{1/2} \\ UCL_\phi &= E(S_n) + 3\sqrt{Var(S_n)} \approx \{E(S_n^2)\}^{1/2} + 3\sqrt{\frac{Var(S_n^2)}{4E(S_n^2)}} \end{aligned} \tag{3.14}$$

Since ϕ is, in general, unknown, we can replace ϕ by its estimator for obtaining the feasible control limits.

4 Control Chart for Road Pavement Data

Griffiths *et al.* (2003) considered data on corse thickness of road pavement of the Barton Highway (linking Yass and Canberra), Australia collected on a two dimensional grid and constructed mean chart with modified control limits keeping the correlation structure among observations in view. They modeled the correlation structures in observations using AR(1) process in one direction and ignored the autocorrelation across the pavement. For numerical illustration purpose, we also consider the same set of data provided by Jim Ollis for 2185 m to 9120 m along three strings namely left pavement (LPAV), center pavement (CPAV) and right pavement (RPAV). For detailed illustration of data one may refer to Griffiths *et al.* (2003). The control limits for mean chart and standard deviation (SD) chart are obtained considering autocorrelation structure, modeled using AR(1) process in both the directions. The control limits are evaluated and charts are constructed using R language. The observations are divided into 9×3 sampling grids. For relative distance $r = 0.76$ (see Griffiths *el al.*, 2003, p. 137), using (2.1) and (2.3), the estimators of autoregressive parameters ϕ_1 and ϕ_2 are $\hat{\phi}_1 = 0.8222051$ and $\hat{\phi}_2 = 0.2222104$, so that $s = 10.11468$. The modified control limits for mean chart, calculated using equation (3.5) are $LCL_\phi = -0.0067451$, $CL_\phi = 0.006905175$ and $UCL_\phi = 0.02055545$. However, if we ignore autocorrelation structure and assume observations to be independent, the control limits are $LCL = 0.0034470$, $CL = 0.006905175$ and $UCL = 0.01036329$. Thus modified control limits are 3.947317 times wider than the unmodified control limits.

Figure 1: $\bar{x}$ chart

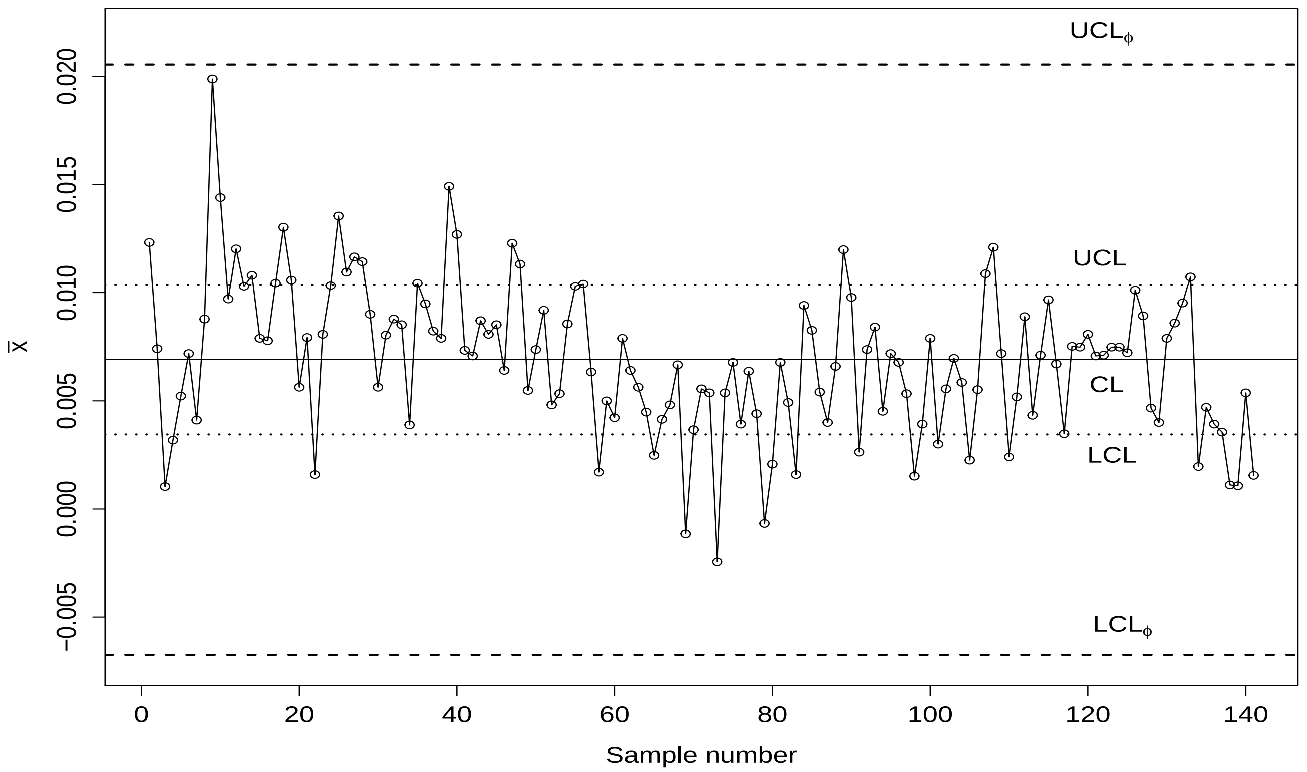

Figure 2: s chart

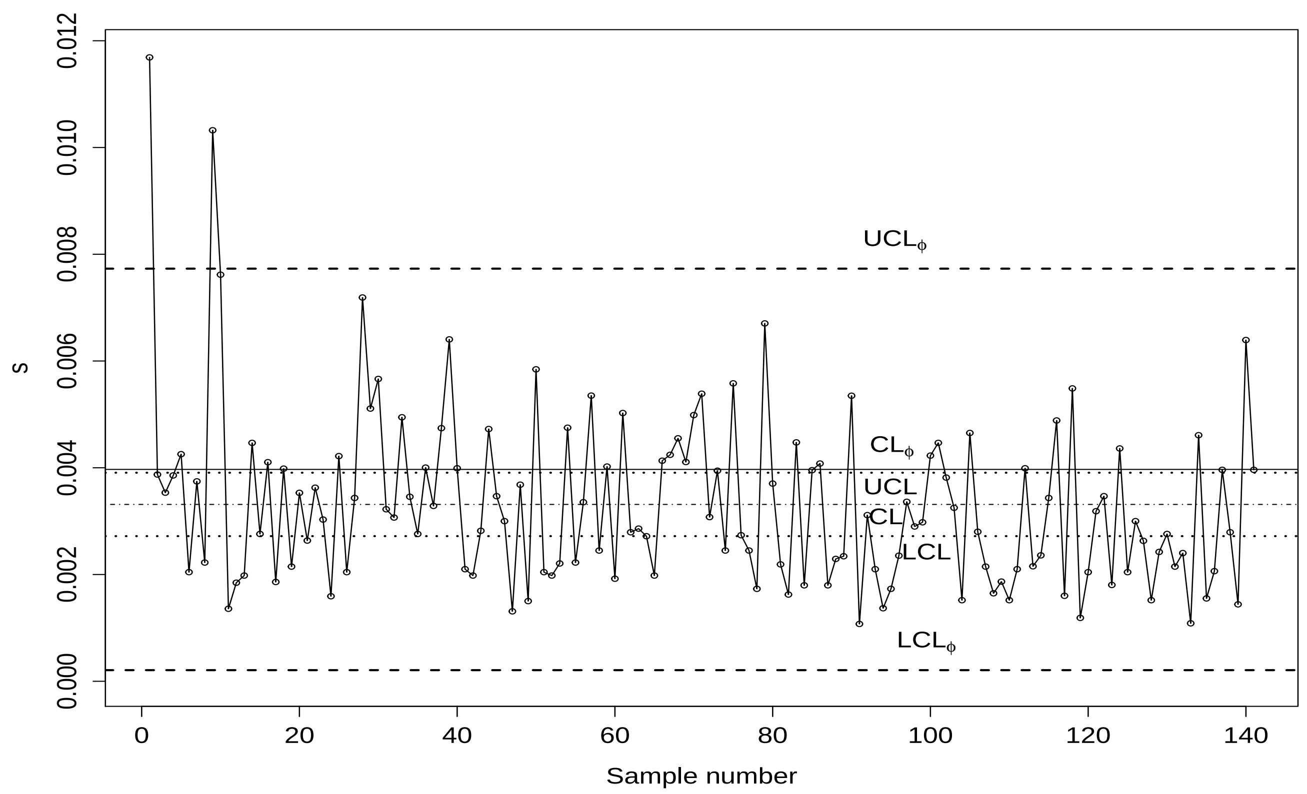

Figure 1 elaborates plot of sample means and leads to the conclusion that the process is in control for means. However, it can be easily verified that if we use unmodified control limits, the process shows false out of control alarm. The modified control limits for s-chart, calculated using equation (3.14)are obtained as $LCL_\phi = 0.0002082855$, $CL_\phi = 0.003969368$ and $UCL_\phi = 0.00773045$. However, if we ignore autocorrelation structure and assume observations to be independent, the control limits are $LCL = 0.002719318$, $CL = 0.003313351$ and $UCL = 0.003907384$. Figure 2 gives plot of s-chart and it has been observed that all the points lie within control bands except sample numbers 1 and 9, which are outside of the upper control limit. It is interesting to observe that the majority of points are not only below the upper control limit but also the central line, leading to the conclusion that sample dispersion is small. As pointed out by Ollis (1997), the two sample numbers (1 and 9) may involve some measurement error(s) leading to such out of control behavior.

We also constructed control charts using 27×3 sampling grids also and observed that the process is in control for both mean and standard deviation. However, the detailed results are not reported here. Our interesting observation is that in general for 27×3 grids, control limits for $\bar{X}$ chart are narrower than control limits for 9×3 grids. However, for standard deviation charts it is other way around, i.e., the control limits for 9×3 grids are narrower. Further, for making means of different grids to be uncorrelated, some gap between the grids is required and that gap depends upon the correlation. For instance, if correlation is 0.82, a gap of twelve rows between two grids reduces the correlation between nearest cells of two grids to 0.09242, which may be considered as negligible.

5 Concluding Remarks

While constructing control charts for observations on two dimensions, we observe that if autocorrelation structure has not been taken into account, the mean and standard deviation control chart may lead to false out of control alarms. The autocorrelation structure may be modeled through considering AR(1) process in both the directions. However, instead of modeling autocorrelation structure in both the directions with two separate processes (or three in case of three dimensional grid), alternatively spatial autoregressive (SAR) model defined over two dimensional or three dimensional grids can be used. The results of the paper may be extended if one gets autocorrelated observations on three dimensions.When process is out of control, the average run length (ARL) properties of the control charts and the behavior of ARL, for different values of ϕ and σ^2 are under investigation.

Acknowledgements

We are very grateful to David Griffiths and John Rayner for their valuable comments. We are also thankful to the editor and three anonymous referees for their constructive suggestions.

References

[1] Capizzi, G. and Massrotto, G. (2008), Practical design of generalized likelihood ratio control charts for autocorrelation data, *Technometrics*, 50(3), 357-370.

[2] Griffiths, D., Gulati, C.M. and Ollis, J. (2003), Statistical control for road pavements, *Australian and New Zealand Journal of Statistics*, 45(2), 129-140.

[3] Jiang, W. (2001), Average run length computation of ARMA charts for stationary processes, *Communications in Statistics, Computation and Simulations*, 30(3), 699-716.

[4] Montgomery, D.C. and Mastrangelo, C.M. (1991), Some statistical process control methods for autocorrelated data, *Journal of Quality Technology*, 23, 179-193.

[5] Ollis, J. (1997), *Statistical Tolerances for Concrete Road Pavement Surfaces*, Masters dissertation, University of Wollongong.

[6] Porter, L.J. and R. Caulcutt (1992), Control Charts Design: A Review of Standard Practice. *Quality and Reliability International*, 8, 113-122.

[7] R Core Team (2014). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. URL http://www.R-project.org/.

[8] Roes, K.C.B. and R.J.M.M. Does (1995), Shewhart-type charts in nonstandard situations,*Technometrics*, vol. 37, pp. 15-24

[9] Shewart, W. (1931), Economic Control of Quality of Manufactured Product, New York: Van Nostrand.

[10] Sparks, R. (2000), CUSUM charts for AR(1) data: Are they worth the effort? *Australian and New Zealand Journal of Statistics*, 42(1), 25-42.

[11] Sparks, R. and J. Ollis (2001), Monitoring Pavement Construction Process, *Advances in Statistics, Combinatorics and Related Areas*, eds. C. Gulati, Yan-Xia Lin, S. Mishra and J. Rayner, 294-308.

[12] Wieringa, J.E. (1999), Statistical Process Control for Serially Correlated Data. Ph.D. Thesis University of Groningen.

[13] Zhang, N.F. (1998a), A statistical control chart for stationary process data, *Technometrics*, 40, 24-38.

[14] Zhang, N.F. (1998b), Estimating process capability indexes for autocorrelated data, *Journal of Applied Statistics*, 25(4), 559-574.

[15] Zhang, N.F. (2002), Statistical process monitoring for autocorrelated data *Advances in Statistics, Combinatorics and Related Areas*, eds. C. Gulati, Yan-Xia Lin, S. Mishra and J. Rayner, 383-394.