QUALITY AND RELIABILITY ENGINEERING INTERNATIONAL
Qual. Reliab. Engng. Int. 2005; **21**:633–653
Published online 10 March 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/qre.680

Research

A General Framework for Constructing Control Charts

Alan M. Polansky*,†
Division of Statistics, Northern Illinois University, De Kalb, IL 60115-2854, U.S.A.

A general framework for the construction of control charts is presented. The method is based on using the density of the sample subgroup statistic as a measure of how unusual newly observed subgroups are. This methodology includes, as special cases, many common control chart techniques. The method is also easily applied to multivariate and multimodal situations. A non-parametric control chart is implemented by estimating the density of the sample subgroup statistic using a kernel estimator of the bootstrap distribution of the observed subgroup statistics. Several examples of the method are presented in the parametric and non-parametric situations. The potential performance of the non-parametric version of the method is demonstrated through an empirical study of average run length properties. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: bootstrap; kernel density estimation; multivariate processes; non-parametric; average run length

1. INTRODUCTION

Consider a manufacturing process that produces items with a d-dimensional quality characteristic X that follows a continuous d-dimensional density f_θ where $\theta \in \mathbb{R}^p$. Let $\psi = D(\theta)$, where $D : \mathbb{R}^p \to \mathbb{R}^m$ for $m \leq p$ and ψ is some characteristic of the process that is to be monitored. If $X_1, X_2, \ldots$ denote the quality characteristics of a sequence of items produced by the process, then the process is considered to be stable, or in control, if these quality characteristics are mutually independent and all follow the density f_θ. As a consequence, the corresponding characteristic ψ is constant for all items from a process that is in control.

The stability of such a process is typically monitored with a control chart. See Part II of Montgomery[1] for a complete review of this topic when $m = 1$. For a review of the case when $m > 1$, see Alt[2], Jackson[3], Mason and Young[4] and Chapter 10 of Montgomery[1]. To construct a control chart we observe k sub-samples of items periodically sampled from the process. In this paper we assume that each sub-sample is of a constant size n. Slight modifications of our method extend it to the case of sub-samples of different sizes. Also, in practice we might not necessarily restrict the control chart to just k sub-samples, but would monitor the process until an out-of-control condition arises. For our current presentation, however, the case of observing k sub-samples of equal size will suffice. Let X_{ij} be the quality characteristic measured on the jth observed item from the ith sub-sample for $i = 1, \ldots, k$ and $j = 1, \ldots, n$. To monitor the process an estimate of ψ is computed using each sub-sample. Let $\hat{\psi}_i$ denote the sample estimate of ψ computed using sub-sample i. Let α be a specified false-alarm probability for the control chart. This is the probability that a single sub-sample will cause the chart

*Correspondence to: Alan M. Polansky, Division of Statistics, Northern Illinois University, De Kalb, IL 60115-2854, U.S.A.
†E-mail: polansky@math.niu.edu

Received 9 July 2003
Revised 20 December 2003

to indicate that the process is out of control when no change in the process has taken place. The typical approach to constructing a control chart requires the construction of a control region $\mathcal{C} \subset \mathbb{R}^m$, which is chosen so that

$$P(\hat{\psi}_i \in \mathcal{C} \mid X_{ij} \text{ is an independent and identically distributed sample from } f_\theta) = 1 - \alpha$$

For such a chart, the process is considered to be in control at subgroup i if $\hat{\psi}_i \in \mathcal{C}$, otherwise the process is considered to be out of control. If $\hat{\psi}_i$ is a distribution-free statistic (see Chapter 2 of Randles and Wolfe[5]), then $\mathcal{C}$ can be computed explicitly and independently of f_θ as long as the sampling distribution of $\hat{\psi}_i$ is known and can be computed. For an example of this type of chart see Alloway and Raghavachari[6]. Usually, however, the region $\mathcal{C}$ will depend on the unknown parameter vector θ and possibly on the exact form of f_θ. In this case we can parametrically model f_θ and compute the control region in terms of f_θ. Hence $\mathcal{C} = \mathcal{C}(f_\theta)$. Note that sometimes this region is only asymptotically correct, but is a good approximation of what the finite sample size region would be. If θ is unknown then we can estimate the control region with $\hat{\mathcal{C}} = \mathcal{C}(f_{\hat{\theta}})$, where $\hat{\theta}$ is some reasonable estimate of θ based either on a training sample, or on the currently observed sample. This is the typical approach to constructing a control chart where the usual parametric model for f_θ is the multivariate normal density. It is well known that some charts constructed in this manner are robust to the assumption about the parametric form of f_θ, while others are not[7–9]. When the chart is not robust to misspecification of the parametric model, a more flexible parametric model (see, for example, Chou *et al.*[10]) or bootstrapping is used to estimate the control region. We use the latter method to implement our proposed non-parametric approach. For another recent application of the bootstrap in constructing control charts see Nichols and Padgett[11].

In this paper we present a general method for constructing control charts. In our approach, we directly use the contours of the density function of $\hat{\psi}_i$ to construct the control region. As we show, this method includes as special cases many of the standard control chart methods in use today, but also easily generalizes to the case where $m > 1$, is valid for any continuous characteristic to be monitored, and also addresses processes that have possibly multi-modal characteristics. Further, we show that studying control charts of this form reveals a simple extension to the non-parametric situation.

We begin in the next section by developing the method in the parametric setting. This is followed by several examples. We next present a non-parametric approach to the method and again explore several examples. The finite sample properties of the non-parametric method are then studied empirically. We conclude with some general comments about the method and the potential for further research in this area.

2. DENSITY CONTROL CHARTS

In this section we present the general method for constructing control charts. Our method is similar to the technique presented on p. 160 of Hall[12] for the construction of multivariate bootstrap confidence sets. Note that, for the purposes of monitoring a process, as long as the characteristic of the output from the process is not unusual given f_θ, then we have no evidence that the process is out of control. On the other hand, if given f_θ, we can conclude that the characteristic of the output from the process is highly unusual then we can conclude that some aspect of the process is out of control. The method we use to detect unusual output from the process is to compare the observed subgroup statistics $\hat{\psi}_1, \ldots, \hat{\psi}_k$ to the density of $\hat{\psi}_i$ calculated under the assumption that the process is in control. The comparison is implemented using what are called 'level-sets' of the density[12].

To motivate the method let $h(x)$ be the density of $\hat{\psi}_i$ for a subsample of size n when the process is in control with the quality characteristics following the density f_θ. Let $\mathcal{S}$ be the support of $h(x)$ and define $\mathcal{I}(c) = \{x \in \mathcal{S} : h(x) \geq c\} \subset \mathcal{S}$ and $\mathcal{O}(c) = \{x \in \mathcal{S} : h(x) < c\} \subset \mathcal{S}$ for an arbitrary positive constant c. Note that $\mathcal{S} = \mathcal{I}(c) \cup \mathcal{O}(c)$ for every $c > 0$. Also, for the selected value of c define

$$\mathcal{A}(c) = \int_{\mathcal{O}(c)} h(x)\,\mathrm{d}x = P(\hat{\psi}_i \in \mathcal{O}(c) \mid X_{ij} \text{ is an independent and identically distributed sample from } f_\theta)$$

We are now able to construct a control chart for the process based on the sample characteristics $\hat{\psi}_1, \hat{\psi}_2, \ldots, \hat{\psi}_k$. Let c_α be the solution of $\mathcal{A}(c) = \alpha$ with respect to c, where α is the desired false-alarm probability for the chart.

For the solution c_α to exist we need to assume that $h(x)$ is an absolutely continuous and almost everywhere differentiable density such that for every $\epsilon > 0$ there exists a value $x_\epsilon \in \mathbb{R}^m$ such that $h(y) \leq \epsilon$ for every y in a neighborhood of radius δ of x_ϵ for some $\delta > 0$. Such assumptions exclude, for example, the possibility that $h(x)$ could be a uniform or truncated density. A less restrictive model is possible, but we will not develop it here. However, most reasonable situations will follow this assumption. To construct the control chart we compute the values $h(\hat{\psi}_1), h(\hat{\psi}_2), \ldots, h(\hat{\psi}_k)$. As long as $h(\hat{\psi}_i) \geq c_\alpha$ we conclude that the process is in control. Whenever $h(\hat{\psi}_i) < c_\alpha$ we conclude that the process is out of control. Further, an obvious graphical chart follows by plotting $h(\hat{\psi}_1), h(\hat{\psi}_2), \ldots, h(\hat{\psi}_k)$ versus i. We can also plot the line $c_{0.5}$ on this chart, which will be the analog of a centerline on a standard control chart. This line will usually not be in the center of the chart, but will be located such that there is a one-half probability that $\hat{\psi}_i$ will be above (or below) this line when the process is in control. Note that as long as $\mathcal{A}(c_\alpha) = \alpha$, the false alarm probability of the chart will be α.

In the case when $h(x)$ is unknown, it must be estimated based either on a training sample taken when the process is known to be in control, or on the current sample. If a parametric form for $h(x)$ is known up to a finite set of unknown parameters, say $\Lambda = (\lambda_1\ \lambda_2 \cdots \lambda_q)'$, then Λ can be estimated based on either sample. Denote this estimate by $\hat{\Lambda}$. The limit c_α can then be estimated by solving $\hat{\mathcal{A}}(c) = \alpha$ where

$$\hat{\mathcal{A}}(c) = \int_{\hat{\mathcal{O}}(c)} \hat{h}(x)\, \mathrm{d}x$$
$$\hat{\mathcal{O}}(c) = \{x \in \mathcal{S} : \hat{h}(x) < c\}$$

and $\hat{h}(x)$ is the parametric form of $h(x)$ with $\Lambda = \hat{\Lambda}$. Alternatively, it is often possible to solve $\mathcal{A}(c) = \alpha$ for a solution c_α that is a function of Λ. Hence $c_\alpha = c_\alpha(\Lambda)$. In this case one can estimate c_α with $\hat{c}_\alpha = c_\alpha(\hat{\Lambda})$. In either case we compare $\hat{h}(\hat{\psi}_i)$ with $\hat{c}_\alpha$ to assess whether the process is in control. Depending on the form of $\hat{\psi}_i$ and the parametric model for h, this chart may only have an approximate false-alarm probability of α. When no parametric form for $h(x)$ is available, a non-parametric procedure for estimating h using smoothing techniques and the bootstrap may be implemented. We outline this procedure after the following examples.

3. EXAMPLES

3.1. Parametric normal $\bar{X}$ charts

Consider the case when f_θ is a univariate normal distribution with mean μ and standard deviation σ. Here we take $\theta' = (\mu\ \ \sigma)$. To monitor the process location we use $\psi = D(\theta) = \mu$ with the obvious estimate $\hat{\psi}_i = \bar{X}_i = n^{-1} \sum_{j=1}^{n} X_{ij}$, for $i = 1, \ldots, k$. Standard theory gives that $h(x)$ is a normal distribution with mean μ and standard deviation $n^{-1/2}\sigma$. Solving $\mathcal{A}(c) = \alpha$ for c yields

$$c_\alpha = \left(\frac{n}{2\pi\sigma^2}\right)^{1/2} \exp(-z_{1-\alpha/2}^2/2)$$

where $z_{1-\alpha/2}$ is the $1-\alpha/2$ percentile of a standard normal distribution. It can be shown that $h(x) \geq c_\alpha$ is equivalent to the condition that $\mu - n^{-1/2} z_{1-\alpha/2}\sigma \leq x \leq \mu - n^{-1/2} z_{1-\alpha/2}\sigma$. We can now estimate μ with $\bar{\bar{X}} = (nk)^{-1} \sum_{i=1}^{k} \sum_{j=1}^{n} X_{ij}$ and σ either with $\hat{\sigma} = \bar{S}/C_4$ where

$$C_4 = \left(\frac{2}{n-1}\right)^{1/2} \frac{\Gamma(n/2)}{\Gamma[(n-1)/2]}$$

or $\hat{\sigma} = \bar{R}/D_2$ where D_2 is given in a table in Appendix VI of Montgomery[1]. Both of these estimates are unbiased under the normality assumption. Hence, this chart consists of concluding that the process is in control at subgroup i if $\bar{\bar{X}} - n^{-1/2} z_{\alpha/2}\hat{\sigma} \leq \bar{X}_i \leq \bar{\bar{X}} - n^{-1/2} z_{\alpha/2}\hat{\sigma}$. This is the standard $\bar{X}$ chart for monitoring process location.

The proposed method also reduces to several other methods for monitoring the univariate location of a process. If the parametric normal assumption is relaxed, then the chart presented above is approximate and is generally justified using the Central Limit Theorem. A better approximation may be available using the bootstrap method which we outline later in this paper. We can also change the form of $\hat{\psi}_i$ to be a Hodges–Lehmann estimator of location and retain the assumption that f_θ is symmetric to produce the chart proposed by Alloway and Raghavachari[6]. If a chart with robust limits is desired, we can simply change the estimates of μ and σ to be those recommended by Langenberg and Iglewicz[13]. Employing an adjustment for the skewness of f_θ results in the chart proposed by Bai and Choi[14].

3.2. *Individuals chart for a normal mixture*

Consider the situation where several machines independently produce identical items which are combined before inspection takes place. Even if each machine produces items that have a quality characteristic that is normal, it is reasonable to model the output from each machine with different means and standard deviations. In this case the quality characteristic for an item sampled from the combined set of items will have a normal mixture distribution, possibly with several separated modes. To simplify the example consider the case of two machines (1 and 2) which produce the same number of items. In this case

$$f_\theta(x) = \frac{1}{2}\sum_{w=1}^{2}(2\pi\sigma_w^2)^{-1/2}\exp[-(x-\mu_w)^2/(2\sigma_w^2)]$$

where $\theta = (\mu_1 \;\; \mu_2 \;\; \sigma_1 \;\; \sigma_2)'$, μ_w is the mean for machine w, and σ_w is the standard deviation for machine w. A reasonable characteristic of interest in this case is the location $\psi = (\mu_1 + \mu_2)/2$ which can be estimated with $\hat{\psi}_i = \bar{X}_i$. In practice one could use a training sample, or the combined sub-samples, to estimate $h(x)$ with an appropriate methodology and solve for c_α. We will not emphasize these methods in this example, but rather explore what the control regions look like in two particular situations.

If the two machines have distributions that are fairly close then the density of the combined items is similar to the density presented in Figure 1. The horizontal dashed line corresponds to a selected value of c_α. We can equivalently calculate upper and lower control limits for $\hat{\psi}_i = \bar{X}_i$ as in the previous example, although it is unlikely that these limits will have a closed form. These limits are indicated by the vertical dotted lines on the plot.

If the two machines have distributions that are more separated, then the density of the combined items is similar to the density presented in Figure 2. Once again the horizontal dashed line corresponds to a selected value of c_α and the vertical dotted lines represent equivalent limits of the control region on $\bar{X}_i$. Note that in this case the corresponding control region for $\hat{\psi}_i = \bar{X}_i$ is the union of two disconnected sets. Hence, we consider the process to be in control if $-4.66 \le \bar{X}_i \le -1.33$ or if $1.33 \le \bar{X}_i \le 4.66$. While this type of control region is not standard, if the purpose of a control chart is to detect unusual output from the process, certainly a point between the two distributions in the mixture would be substantial evidence that the process is no longer in control. While such a situation might imply quality improvement, new control limits would need to be established. While our proposed chart would be able to detect such a change, the standard Shewhart chart for location would not detect such a change since its control region is always a single connected interval.

3.3. *Bivariate location control charts*

Consider the case where X_{ij} is a 2×1 vector of the form $X_{ij} = (W_{ij1} \;\; W_{ij2})'$ and f_θ is a bivariate normal distribution with $\theta = (\mu_1 \;\; \mu_2 \;\; \sigma_1 \;\; \sigma_2 \;\; \rho)'$, where $\mu_q = E(W_{ijq})$, $\sigma_q^2 = E[(W_{ijq} - \mu_q)^2]$, and ρ is the correlation between W_{ij1} and W_{ij2}. To monitor the location of this process we can take $\psi = (\mu_1 \;\; \mu_2)'$ so that it is reasonable to take $\hat{\psi}_i = \bar{X}_i = n^{-1}\sum_{j=1}^{n} X_{ij}$, the vector sample mean of the subgroup. Under the assumption of normality we have that $h(x)$ is a bivariate normal distribution with $\theta = (\mu_1 \;\; \mu_2 \;\; \sigma_1/\sqrt{n} \;\; \sigma_2/\sqrt{n} \;\; \rho)'$. In this case c_α is a

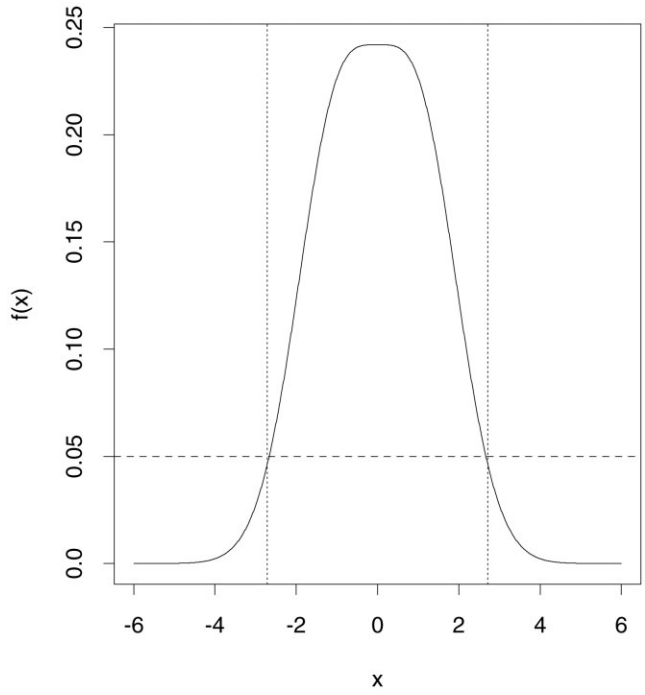

Figure 1. First example density for the mixture process in the second example when $n = 1$. The dashed horizontal line is the density control limit c_α and the dotted vertical lines correspond to the limits of the control regions for an equivalent chart for $\hat{\psi}_i = \bar{X}_i$

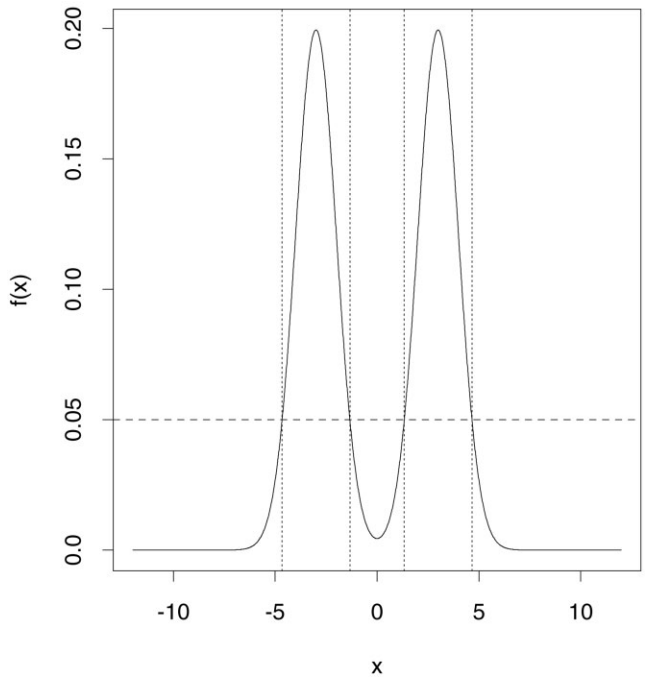

Figure 2. Second example density for the mixture process in the second example when $n = 1$. The dashed horizontal line is the density control limit c_α and the dotted vertical lines correspond to the limits of the control regions for an equivalent chart for $\hat{\psi}_i = \bar{X}_i$

constant such that

$$\int_{\mathcal{I}(c_\alpha)} h(x)\,\mathrm{d}x = 1-\alpha$$

where the integral in the equation above is assumed to be integrated over $\mathbb{R}^2$. The function $\mathcal{I}(c_\alpha)$ is defined to be $\mathcal{I}(c_\alpha) = \{x \in \mathbb{R}^2 : (x-\mu)'\Sigma^{-1}(x-\mu) \le -2\ln(2\pi|\Sigma|^{1/2}c_\alpha)\}$ where $\mu = (\mu_1 \;\; \mu_2)'$ and

$$\Sigma = \begin{bmatrix} \sigma_1^2/n & \sigma_1\sigma_2\rho/n \\ \sigma_1\sigma_2\rho/n & \sigma_2^2/n \end{bmatrix}$$

It is well known that $-2\ln(2\pi|\Sigma|^{1/2}c_\alpha) = \chi^2_{2,1-\alpha}$, where $\chi^2_{\nu,\xi}$ is the ξth percentile of a chi-squared distribution with ν degrees of freedom. This yields

$$c_\alpha = (2\pi)^{-1}|\Sigma|^{-1/2}\exp(-\chi^2_{2,1-\alpha}/2)$$

If μ and Σ are unknown, then we can estimate them on the basis of a large training sample. Denote these estimates by $\hat{\mu}$ and $\hat{\Sigma}$, respectively. Hence, c_α can be estimated by

$$c_\alpha = (2\pi)^{-1}|\hat{\Sigma}|^{-1/2}\exp(-\chi^2_{2,1-\alpha}/2)$$

and $h(x)$ can be estimated with a bivariate normal density with mean vector $\hat{\mu}$ and covariance matrix $\hat{\Sigma}$. Note that this is equivalent to the usual phase-II T^2 chart (see p. 111 of Alt[2]). The limit $\hat{c}_\alpha$ is technically not the correct limit since μ and Σ are estimated and not known. However, a large enough training sample may provide a reasonable approximation.

A better approach relies on a training sample that is independent of the current observations. In this case (see, for example, p. 119 of Alt[2])

$$(\hat{\psi}_i - \hat{\mu})'\hat{\Sigma}^{-1}(\hat{\psi}_i - \hat{\mu}) \sim \frac{2(q+1)(n-1)}{n(nq-q-1)}F_{2,nq-q-1}$$

where q is the number of subgroups observed in the training sample and $F_{v,w}$ represents an F distribution with v and w degrees of freedom. Hence, we have that

$$c_\alpha = (2\pi)^{-1}|\Sigma|^{-1/2}\exp\left[-\frac{(q+1)(n-1)}{n(nq-q-1)}F_{2,nq-q-1,1-\alpha}\right]$$

where $F_{2,nq-q-1,\xi}$ is the ξth percentile of the $F_{2,nq-q-1}$ distribution. This provides us with an exact density control chart in the bivariate normal situation. Note that this chart is equivalent to the usual phase-I T^2 chart.

4. A NON-PARAMETRIC APPROACH

The form of the control charts presented in the previous section suggests a simple method for implementing non-parametric control charts. Non-parametric control charts are useful when the practitioner does not want to model f_θ parametrically either because of a lack of information about f_θ or as a secondary check on the conclusions of a parametric chart.

To construct a non-parametric version of our proposed chart we must first compute a non-parametric estimate of $h(x)$. The method we suggest is based on the non-parametric bootstrap. The bootstrap has been implemented by many researchers in the construction of control charts. The performance of many of these charts was recently questioned by Jones and Woodall[15]. Our application of the bootstrap differs slightly from previous methods. Further, as we demonstrate in our empirical study, the technique we use provides reasonable performance. Once the estimate of $h(x)$ is computed we can attempt to construct a control chart based on our estimate

$h(x)$ and the methodology outlined earlier in this paper. However, the bootstrap approach we employ yields a discrete distribution as an estimate of $h(x)$ whereas our density control chart requires a continuous density. To overcome this difficulty we apply a non-parametric smoother to the bootstrap estimate to obtain a continuous non-parametric estimate of $h(x)$. There are numerous methods available for providing such a continuous estimate[16–18]. For simplicity, we emphasize kernel density estimation in this paper. Other methods may be more reliable in certain situations, but will fit into our framework just the same. It is important, however, that the density estimation method produce a continuous or smooth estimate of the unknown density. For a review of kernel density estimation see Chapters 2 and 4 of Wand and Jones[19].

4.1. *Univariate case*

To develop the univariate case consider a training sample of b subgroups of size n taken when the process is known to be in control. One could use the currently observed data as the training sample, but the analysis would be less precise since the process may not be currently in control. The training sample itself can be multivariate, but in this section we assume that the dimension of $\hat{\psi}_i$ is $m = 1$.

To produce a non-parametric estimate of $h(x)$ we use the resampling technique popularized by Efron[20]. Denote the combined training sample by $T_{11}, \ldots, T_{bn}$ and construct B resamples of size n from $T_{11}, \ldots, T_{bn}$. Each resample is constructed by randomly drawing n times from $T_{11}, \ldots, T_{bn}$ with replacement. For each resample compute $\hat{\psi}$. Denote these estimates by $\hat{\psi}_1^*, \ldots, \hat{\psi}_B^*$.

This resampling method implicitly assumes that the training sample is an independent and identically distributed sample from f_θ. In cases where this assumption may not be valid, such as when the current observations are being used as the training sample, a different resampling approach is required. The subgroup bootstrap approach[21] is one such procedure that can be used with our proposed method to generate $\hat{\psi}_1^*, \ldots, \hat{\psi}_B^*$. In relation to other bootstrap control chart techniques, our method is similar in approach to the method of Bajgier[22] and Liu and Tang[23] in the independent case, except that we use a smooth estimate for the estimated sampling distribution. If the subgroup sizes are large enough we suggest using a studentized version of these methods that generates resampled statistics of the form $\hat{\tau}_i^* = (\hat{\psi}_i^* - \hat{\psi})/\hat{\sigma}_i^*$, where $\hat{\sigma}_i^*$ is a standard error estimate of $\hat{\psi}_i^*$ based on an analytic form, jackknife or bootstrap calculations and $\hat{\psi}$ is an estimate of ψ based on the entire training sample. Such a form takes advantage of certain asymptotic properties that produce a more accurate estimate of the sampling distribution $h(x)$. See Chapter 3 of Hall[12] for more information on these methods. The studentized method should not be used when the subgroup size is small as the behavior of $\hat{\sigma}_i^*$ in the denominator of $\hat{\tau}_i^*$ can lead to instability. In this case the smoothed studentized version of the bootstrap developed by Polansky[24] may be more appropriate.

To construct a continuous non-parametric estimate of $h(x)$ we compute a kernel density estimate of the form

$$\hat{h}_t(x) = \frac{1}{Bt} \sum_{i=1}^{B} K\left(\frac{\hat{\psi}_i^* - x}{t}\right)$$

for all $x \in \mathbb{R}$, where K is an absolutely continuous and symmetric density called a kernel function and t is a tuning parameter called a bandwidth. The selection of the function K is generally considered to be unimportant asymptotically[18], and hence we assume that K is a standard normal density for the remainder of this paper. Note that the kernel function should follow the same assumptions as imposed on $h(x)$ outlined earlier in this paper. This will ensure that the kernel estimate also follows the same assumptions. The selection of the tuning parameter t is crucial in density estimation. In our application we will be using the estimated density function as a measure of 'data depth' and hence the optimal estimation of the density will be less crucial, but still very important. There have been many methods developed for choosing a reasonable value of t (see Chapter 3 of Wand and Jones[19]). In our case we use the two-stage plug-in method which is outlined in Sheather and Jones[25]. This method has shown great promise in the density estimation literature and is relatively simple to compute. The method we present here is outlined on p. 72 of Wand and Jones[19] and proceeds as follows.

(1) Compute the estimate

$$\hat{C}_1 = \hat{\sigma}\left[\frac{-64\pi^{1/2}K^{(6)}(0)}{105\mu_2(K)B}\right]^{1/9}$$

where $K^{(6)}$ is the sixth derivative of K,

$$\mu_2(K) = \int x^2 K(x)\,\mathrm{d}x$$

and $\hat{\sigma}$ is an estimate of scale. To estimate σ we will use the suggestion on p. 47 of Silverman[26] and use

$$\hat{\sigma} = \min\{S, \mathrm{IQR}/1.349\}$$

where S is the sample standard deviation and IQR is the interquartile range of $\hat{\psi}_1^*, \ldots, \hat{\psi}_B^*$.

(2) Compute the estimate

$$\hat{C}_2 = B^{-2}\hat{C}_1^{-7}\sum_{i=1}^{B}\sum_{j=1}^{B}K^{(6)}[(\hat{\psi}_i^* - \hat{\psi}_j^*)/\hat{C}_1]$$

(3) Compute the estimate

$$\hat{C}_3 = \left[\frac{-2K^{(4)}(0)}{\mu_2(K)\hat{C}_2 B}\right]^{1/7}$$

(4) Compute the estimate

$$\hat{C}_4 = B^{-2}\hat{C}_3^{-5}\sum_{i=1}^{B}\sum_{j=1}^{B}K^{(4)}[(\hat{\psi}_i^* - \hat{\psi}_j^*)/\hat{C}_3]$$

(5) Compute the bandwidth estimate

$$\hat{t} = \left[\frac{R(K)}{\mu_2^2(K)\hat{C}_4 B}\right]^{1/5}$$

where

$$R(K) = \int K^2(x)\,\mathrm{d}x$$

There are arguably more accurate versions of this method that require more computation, but this two-stage method performs reasonably well in practice and is the most common choice. Software for the computation of this bandwidth is available from the author, and also through the S-Plus libraries by Bowman and Azzalini[16] and Wand and Jones[19].

A side effect of smoothing the bootstrap resamples using the kernel estimate is that the variance of the distribution $\hat{h}_t(x)$ will be larger than the bootstrap estimate of the variance of $h(x)$. The bootstrap estimate of the standard error of $\hat{\psi}_i$ is

$$\hat{\sigma}^* = B^{-1}\sum_{i=1}^{B}(\hat{\psi}_i^* - \bar{\psi}^*)^2$$

where $\bar{\psi}^* = B^{-1} \sum_{i=1}^{B} \hat{\psi}_i^*$. It is reasonable to require that the estimate of $h(x)$ have this same standard deviation. However, it can be shown that

$$\int (x - \bar{\psi}^*)^2 \hat{h}_t(x)\, \mathrm{d}x = B^{-1} \sum_{i=1}^{B} (\hat{\psi}_i^* - \bar{\psi}^*)^2 + t^2 = (\hat{\sigma}^*)^2 + t^2$$

To account for this inflation in the variance we will rescale $\hat{\psi}_1^*, \ldots, \hat{\psi}_B^*$ as follows. Assume that $t^2 < (\hat{\sigma}^*)^2$ and define $\tilde{\psi}_i^*$ to be given by

$$\tilde{\psi}_i^* = [(\hat{\sigma}^*)^2 - t^2]^{1/2} (\hat{\psi}_i^* - \bar{\psi}^*)/\hat{\sigma}^* + \bar{\psi}^*$$

Let $\tilde{h}_t(x)$ be defined as

$$\tilde{h}_t(x) = \frac{1}{Bt} \sum_{i=1}^{B} K\left(\frac{\tilde{\psi}_i^* - x}{t}\right)$$

then we have that

$$\int (x - \bar{\psi}^*)^2 \tilde{h}_t(x)\, \mathrm{d}x = (\hat{\sigma}^*)^2$$

Once the estimate of the optimal bandwidth has been computed on the original resampled subgroup statistics and the resampled subgroup statistics have been rescaled, we then take $\tilde{h}_{\hat{t}}(x)$ to be our estimate of $h(x)$. Using this estimate we then calculate $\tilde{c}_\alpha$ to be the solution to

$$\int_{\tilde{\mathcal{O}}(c)} \tilde{h}_{n,\hat{t}}(x)\, \mathrm{d}x = \alpha$$

with respect to c where $\tilde{\mathcal{O}}(c) = \{x \in \mathbb{R} : \tilde{h}_{\hat{t}}(x) < c\}$. The control chart is then constructed by plotting $\tilde{h}_{\hat{t}}(\hat{\psi}_1), \ldots, \tilde{h}_{\hat{t}}(\hat{\psi}_k)$ versus i and using $\tilde{c}_\alpha$ as the control limit. We can similarly compute the centerline $\tilde{c}_{0.5}$. The above chart can also be used to numerically construct equivalent limits for $\hat{\psi}_1, \ldots, \hat{\psi}_k$, if desired.

4.2. *Example*

As an example we consider the set of measurements of the inside diameter of forged piston rings (measured in millimeters) studied by Montgomery[1] on pp. 213–215. The first 25 subgroups of size 5 from this process where taken when the process is in control and will be used as the training sample. A second set of 15 subgroups of size 5 were taken later in the process and will serve as our current observations. We first construct a density chart for location based on the subgroup means. We pooled the subgroups in the training sample and constructed 1000 resampled subgroups of size $n = 5$, which yielded 1000 $\hat{\psi}_i^*$ values. The two-stage bandwidth estimate was computed on these 1000 subgroup means resulting in an estimated bandwidth of $\hat{t} = 0.001\,215$. From the 1000 resamples we found $\hat{\sigma}^* = 0.000\,020\,61$ so that the rescaling factor is $[(\hat{\sigma}^*)^2 - t^2]/\hat{\sigma}^* = 0.9636$. The kernel density estimate $\tilde{h}_{\hat{t}}(x)$ is presented in Figure 3. We set $\alpha = 0.01$ and used numerical integration and a simple bisection root finding routine (see, for example, Chapters 3 and 7 of Conte and de Boor[27]) to find $\hat{c}_{0.01} = 3.339\,980$ and $\hat{c}_{0.5} = 68.930\,93$. The density chart corresponding to the current 15 subgroups is plotted in Figure 4. From the chart one can observe that an out-of-control condition is detected at subgroup number 9. There is also a visible downward trend in the chart indicating the location shift. These conclusions are consistent with those given by Montgomery[1].

We repeated this process to investigate the variability of the process using $\hat{\psi}_i$ as the sample range. The two-stage bandwidth estimated from the 1000 resamples is $\hat{t} = 0.001\,983$ with $\hat{\sigma}^* = 0.000\,082\,12$ so that the rescaling factor is $[(\hat{\sigma}^*)^2 - t^2]/\hat{\sigma}^* = 0.9758$. We used 1000 resamples and $\alpha = 0.01$. The kernel density estimate of $\tilde{h}_{\hat{t}}(x)$ is plotted in Figure 5 and the density chart for the ranges is plotted in Figure 6. The conclusion from this chart is consistent with Montgomery[1] that the variability of the process is in control.

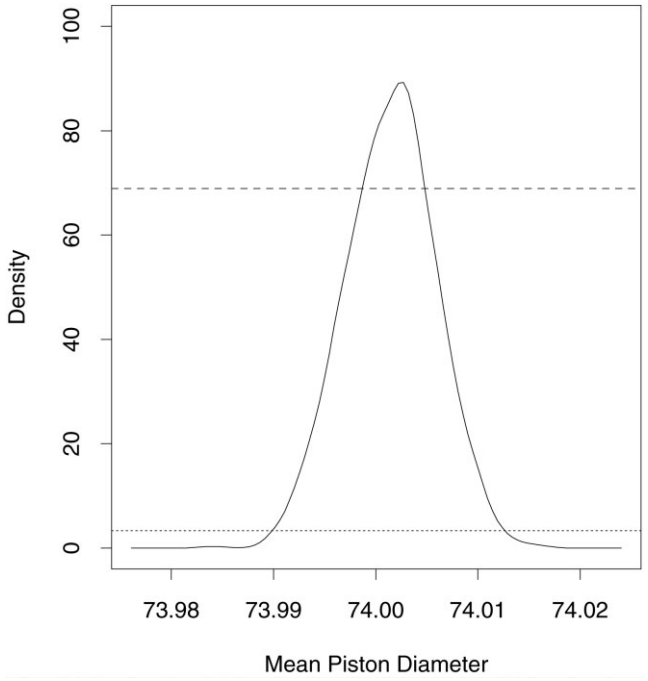

Figure 3. Estimated density $\tilde{h}_t(x)$ for the subgroup mean of the piston ring data. The dotted line is the control limit and the dashed line is the centerline

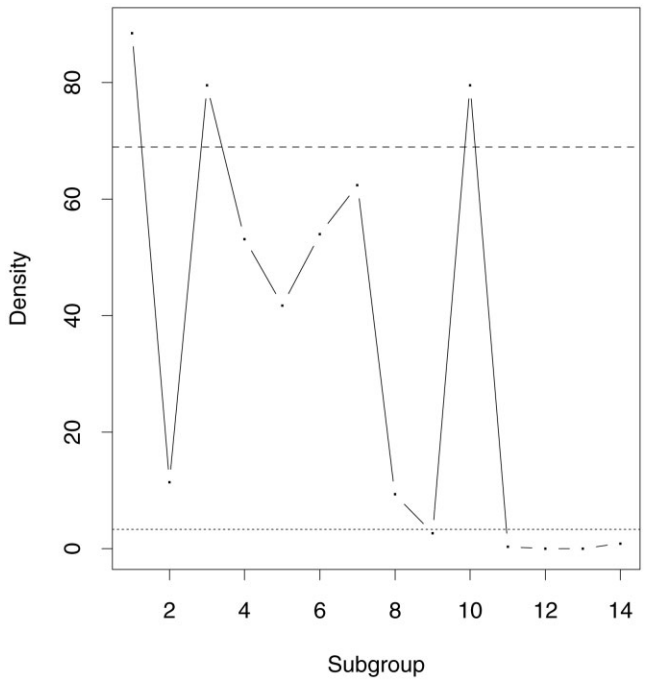

Figure 4. The density control chart for the subgroup mean of the piston ring data. The dotted line is the control limit and the dashed line is the centerline

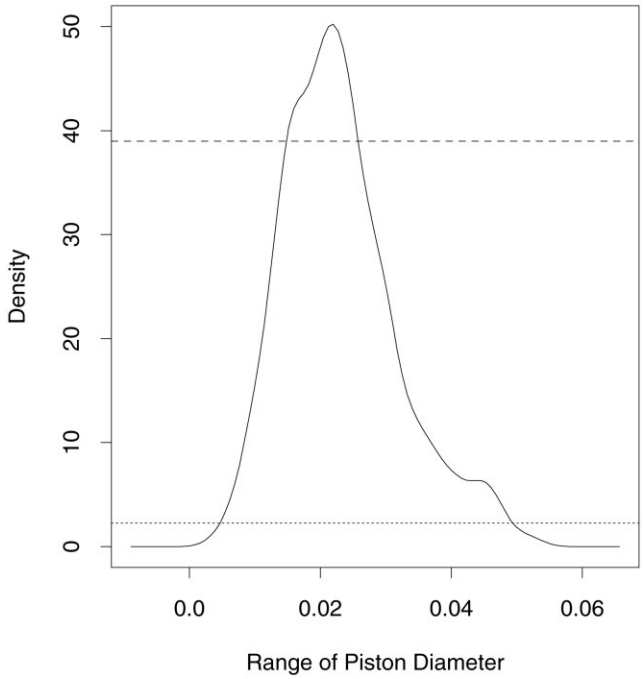

Figure 5. Estimated density $\tilde{h}_{\hat{f}}(x)$ for the subgroup range of the piston ring data. The dotted line is the control limit and the dashed line is the centerline

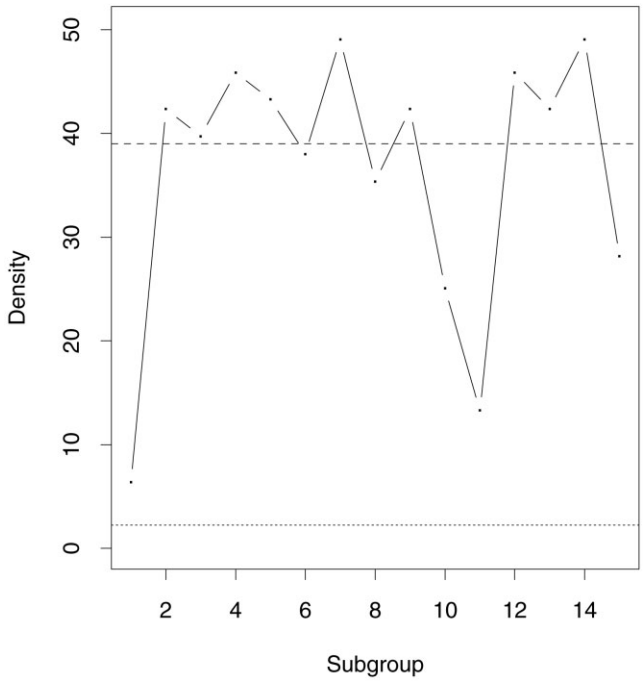

Figure 6. The density control chart for the subgroup range of the piston ring data. The dotted line is the control limit and the dashed line is the centerline

4.3. Multivariate Case

To develop the multivariate case again consider a training sample of b subgroups of size n. From this sample we generate $\hat{\psi}_1^*, \ldots, \hat{\psi}_B^*$ based on the resampling algorithms outlined above. In this case we assume that $m > 1$ so that each $\hat{\psi}_i^*$ is a $m \times 1$ vector and $h(x)$ is an m-dimensional density.

To construct a multivariate kernel estimate of $h(x)$ let $K : \mathbb{R}^m \to \mathbb{R}$ be a continuous function called an m-variate kernel function. Let $x = (x_1 \cdots x_m)'$ be an arbitrary $m \times 1$ vector. The function K is assumed to satisfy the following conditions:

$$\int K(x)\,\mathrm{d}x = 1$$

$$\int xK(x)\,\mathrm{d}x = 0_{m\times 1}$$

where $0_{r\times c}$ is an $r \times c$ matrix with all elements equal to zero, and

$$\int xx'K(x)\,\mathrm{d}x = \mu_2(K)I_{m\times m}$$

where

$$\int \mu_2(K) = x_i^2 K(x)\,\mathrm{d}x$$

is independent of i and $I_{m\times m}$ is a $m \times m$ identity matrix. All of the integrals are assumed to be integrated over the entire space $\mathbb{R}^m$. Given these assumptions, K can be taken to be a multivariate density with mean vector $0_{m\times 1}$ and covariance matrix $\mu_2(K)I_{m\times m}$. Methods for generating K from a univariate density under the assumption of independence or radial symmetry is discussed in Section 4.2 of Wand and Jones[19]. In this paper, we take K to be an m-dimensional normal distribution with mean vector $0_{m\times 1}$ and covariance matrix $I_{m\times m}$. Let T be a symmetric positive definite $m \times m$ matrix called a bandwidth matrix and define

$$K_T(x) = |T|^{-1/2}K(T^{-1/2}x)$$

The m-variate kernel density estimator of $h(x)$ as introduced in its general form by Deheuvels[28] is given by

$$\hat{h}_T(x) = B^{-1}\sum_{i=1}^{B} K_T(x - \hat{\psi}_i^*)$$

for all $x \in \mathbb{R}^m$. The bandwidth matrix T plays the same role as the tuning parameter t in the univariate case. However, bandwidth matrix estimation for the multivariate case is considerably more difficult than in the univariate case. A method similar to that of Sheather and Jones[25] that we recommend for the univariate case has been developed by Wand and Jones[29], but is difficult to implement in practice. For a review of other methods used in the multivariate case see Chapter 6 of Scott[17] or Chapter 4 of Wand and Jones[19].

We use a simple method, which appears to work reasonably well. The method is motivated as follows. Suppose $\hat{\psi}_i$ were to follow an m-dimensional normal distribution with mean vector μ and covariance matrix Σ. If K is an m-dimensional normal distribution with mean vector $0_{m\times 1}$ and covariance matrix $I_{m\times m}$, then the asymptotically optimal bandwidth matrix is given by

$$T_\phi = \left[\frac{4}{m+2}\right]^{2/(m+4)} \Sigma B^{-2/(m+4)}$$

which can be estimated with

$$\hat{T}_\phi = \left[\frac{4}{m+2}\right]^{2/(m+4)} \hat{\Sigma} B^{-2/(m+4)}$$

where $\hat{\Sigma}$ is an estimate of Σ. While this bandwidth matrix estimate is optimal only for the normal density, the method is still flexible enough to adapt to many differently shaped distributions as we show in the examples below. A simpler estimate can be implemented by restricting T to be a diagonal matrix. In many cases there is little loss of efficiency from making this restriction. See Wand and Jones[30] for more information on this restriction.

As in the univariate case, multivariate kernel smoothing will change the covariance structure of the resamples. We rescale the resamples so that $\hat{h}_T(x)$ retains the bootstrap estimate of the covariance structure of $h(x)$. Denote the covariance matrix of the resamples by $\hat{\Sigma}^*$. Then the covariance matrix of $\hat{h}_{\hat{T}_\phi}(x)$ is $\hat{\Sigma}^* + \hat{T}_\phi$. To rescale the resampled random vectors, apply the transformation

$$\tilde{\psi}_i^* = (\hat{\Sigma}^* - \hat{T}_\phi)^{1/2}(\hat{\Sigma}^*)^{-1/2}(\hat{\psi}_i^* - \bar{\psi}^*) + \bar{\psi}^*$$

where $\bar{\psi}^* = B^{-1}\sum_{i=1}^{B}\psi_i^*$. We must assume that $\hat{\Sigma}^* - \hat{T}_\phi$ is a positive definite matrix, otherwise we will be unable to rescale the resampled observations as indicated. Again we define the rescaled kernel estimate by computing the kernel estimate on $\tilde{\psi}_1^*, \ldots, \tilde{\psi}_B^*$, or as

$$\tilde{h}_T(x) = B^{-1}\sum_{i=1}^{B} K_T(x - \tilde{\psi}_i^*)$$

As before, in the examples that follow we use $\tilde{h}_{\hat{T}_\phi}(x)$ as our estimate of $h(x)$. Similarly, we define $\tilde{c}_\alpha$ to be the solution of

$$\int_{\tilde{\mathcal{O}}(c)} \tilde{h}_{\hat{T}_\phi}(x)\,\mathrm{d}x = \alpha$$

with respect to c, where the integral is now taken over $\mathbb{R}^m$ and $\tilde{\mathcal{O}}(c) = \{x \in \mathbb{R}^m : \tilde{h}_{\hat{T}_\phi}(x) < c\}$. We can similarly compute the centerline $\tilde{c}_{0.5}$. The control chart is then constructed by plotting $\tilde{h}_{\hat{T}_\phi}(\hat{\psi}_1), \ldots, \tilde{h}_{\hat{T}_\phi}(\hat{\psi}_k)$ versus i and using $\tilde{c}_\alpha$ as the control limit. One can alternately use the above chart to numerically construct equivalent regions for $\hat{\psi}_1, \ldots, \hat{\psi}_k$ in $\mathbb{R}^m$, but the density chart is much easier to interpret in this case.

In the parametric case, these charts reduce to the multivariate charts used in practice based on the assumption of multivariate normality, where the contours of the parametric density $h(x)$ are used to construct the control region. In the non-parametric setting, we use the same idea except that the contours of the non-parametric kernel estimate $\tilde{h}_{\hat{T}_\phi}(x)$ are used to construct the control region. This is similar to the method by Liu[31]. Indeed, the methods appear to complement one another. For example, one can develop smooth versions of the charts of Liu[31] by replacing the data depth measure with the kernel density estimate used in this paper. Correspondingly, in the density charts presented in this paper, one can replace the kernel estimate with the data depth measure used by Liu[31] to develop a density chart based on multivariate ranks. The measure used by Liu[31] is based on the simplicial depth measure developed by Liu[32].

When the chart indicates an out-of-control condition we need to determine which component, or combination of components, of the multivariate process caused the process to go out of control. Since the density chart essentially reduces the problem of monitoring the multivariate process to a univariate one, it is not possible to extract this information directly from the chart. We suggest that if an out-of-control condition arises, then separate density charts for each component of the process should be examined. If a single component has caused the out-of-control condition, then it should be obvious from one of these charts. Such charts are relatively simple to construct using the multivariate kernel estimate. Let $h_i(x)$ be the marginal density of the ith component of the process and let $x_{[i]}$ denote the $(m-1)$-dimensional column vector that is without the ith component. Then the fact that

$$h_i(x_i) = \int h(x)\,\mathrm{d}x_{[i]}$$

yields a reasonable estimate for $h_i(x_i)$ as

$$\tilde{h}_i(x_i) = \int \tilde{h}_{\hat{T}_\phi}(x)\, \mathrm{d}x_{[i]} = B^{-1} \sum_{i=1}^{B} \int K_{\hat{T}_\phi}(x - \tilde{\psi}_i^*)\, \mathrm{d}x_{[i]} = B^{-1} \sum_{i=1}^{B} K_{i,\hat{T}_\phi}(x - \tilde{\psi}_i^*)$$

where $K_{i,T}$ is the ith marginal distribution of K_T. Hence, we only need to compute the marginal density of the kernel function to compute the marginal density of the multivariate kernel estimate. If no univariate marginal chart is out of control, then we proceed by examining all possible bivariate marginal density charts for the process. These charts are constructed in a similar manner to the univariate charts above. If none of these charts are out of control, then we proceed by examining charts of increasing dimension until we find the charts with the lowest dimension that are out of control. The collection of these charts will indicate what characteristics of the process caused the process to go out of control. Note that it is possible that *none* of the marginal charts will indicate an out-of-control condition. In this case the best conclusion we can make is that all of the process components simultaneously caused the process to go out of control.

4.4. *Example*

We again consider data on the inside diameter measurements of forged piston rings studied on pp. 213–215 of Montgomery[1]. In this case we construct a density chart that simultaneously monitors both location and variability. Here we take $m = 2$ with $\hat{\psi} = (\bar{X}, R)$ where R is the subgroup range. We pooled the subgroups in the training sample and constructed 1000 resampled subgroups of size $n = 5$, which yielded 1000 $\hat{\psi}_i^*$ values. Note that each $\hat{\psi}_i^*$ is a 2×1 vector. The estimated bandwidth matrix computed on these 1000 subgroup was

$$\hat{T}_\phi = \begin{bmatrix} 0.000\,002\,05 & -0.000\,000\,51 \\ -0.000\,000\,51 & 0.000\,008\,89 \end{bmatrix}$$

The resampled statistics where rescaled as described above. A contour plot of the bivariate kernel density estimate $\tilde{h}_{\hat{T}_\phi}(x)$ is presented in Figure 7. We set $\alpha = 0.01$ and used bivariate numerical integration and a bisection root finding routine to find $\hat{c}_{0.01} = 47.69$ and $\hat{c}_{0.50} = 2112.71$. The density chart corresponding to the current 15 subgroups is plotted in Figure 8. Note that, as in Montgomery[1], we detect an out-of-control condition and there is a definite downward trend in the chart. Univariate charts similar to those presented in Figures 4 and 6 would reveal that the change was due to a shift in location. We also computed the equivalent control region for the $\hat{\psi} = (\bar{X}, R)$. This region is plotted in Figure 9, along with the region corresponding top the centerline. The 15 current observations are plotted and numbered on this chart for comparison.

4.5. *Example*

For a second example we consider the bivariate chemical data given in Table 2 of Tracy *et al.*[33]. The data are used in a slightly different context in this paper, but provide an interesting set of bivariate data that are correlated. Define each observation as a 2×1 vector (X_j, Y_j), where X_j is a unit lab measurement and Y_j is a main lab measurement at time j. With this example we are interested in monitoring the bivariate location so that we will take $\hat{\psi}_i = (\bar{X}_i, \bar{Y}_i)$, where $\bar{X}_i$ is the subgroup mean of the unit lab measurements and $\bar{Y}_i$ is the subgroup mean of the main lab measurements for subgroup i. As in Tracy *et al.*[33] we treat the first 15 observations as a training sample. The remaining four observations are combined into a single subgroup so that $n = 4$. From the training sample we constructed 1000 resampled subgroups of size $n = 4$, which yielded 1000 $\hat{\psi}_i^*$ values. Note again that each $\hat{\psi}_i^*$ is a 2×1 vector. The estimated bandwidth matrix was computed on these 1000 subgroup means resulting in

$$\hat{T}_\phi = \begin{bmatrix} 0.003\,609\,39 & 0.000\,848\,35 \\ 0.000\,848\,35 & 0.000\,594\,51 \end{bmatrix}$$

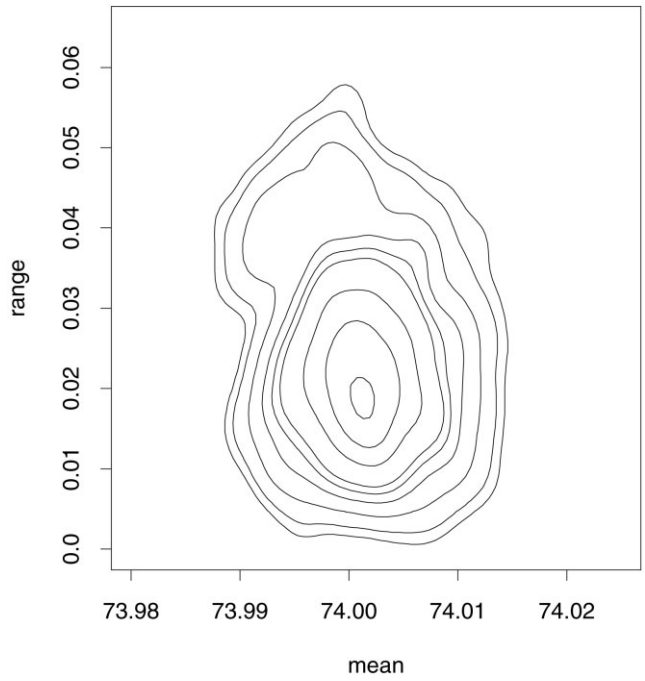

Figure 7. Density contours of the bivariate kernel estimate $\tilde{h}_{\hat{T}_\phi}(x)$ of the sampling distribution of $\hat{\psi}_i = (\bar{X}_i, R_i)$, for the piston ring data

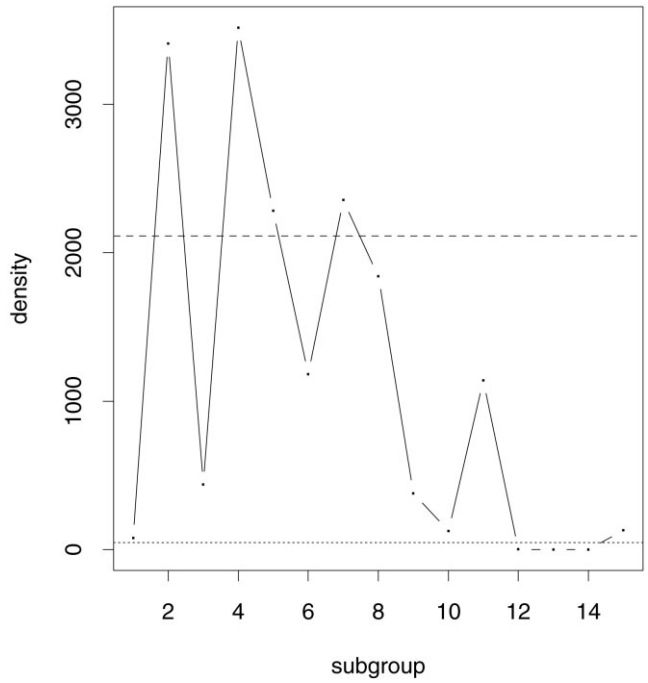

Figure 8. Density control chart for the bivariate process of $\hat{\psi}_i = (\bar{X}_i, R_i)'$ for the piston ring data. The dotted line is the control limit and the dashed line is the centerline

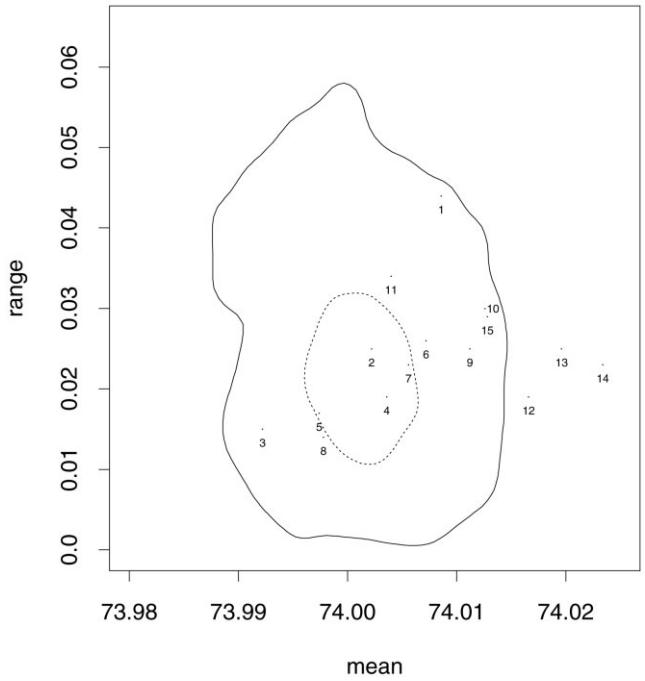

Figure 9. Equivalent control region of the density chart for the bivariate process of $\hat{\psi}_i = (\bar{X}_i, R_i)'$ for the piston ring data. The area inside the solid curve is the control region. The dashed curve is the contour corresponding to the centerline. The numbered points correspond to the current 15 observations

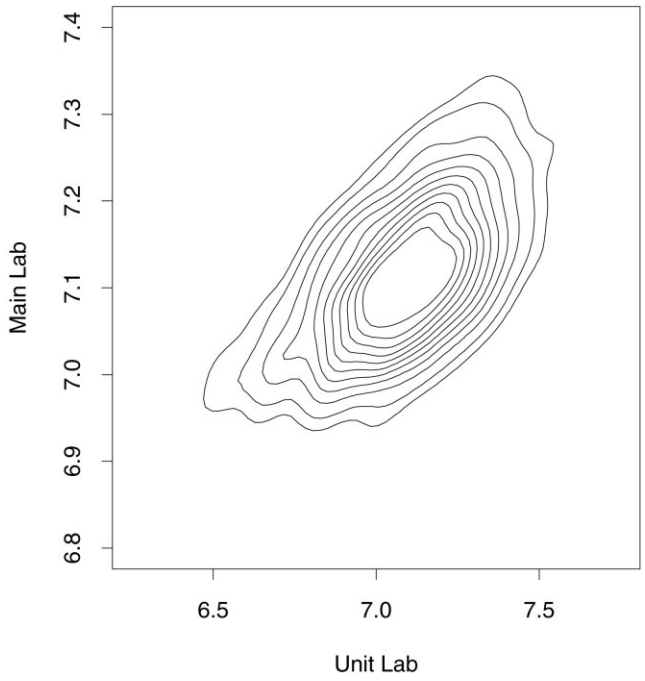

Figure 10. Density contours of the bivariate kernel estimate $\tilde{h}_{\hat{T}_\phi}(x)$ of the sampling distribution of $\hat{\psi}_i = (\bar{X}_i, \bar{Y}_i)$, for the lab measurement data

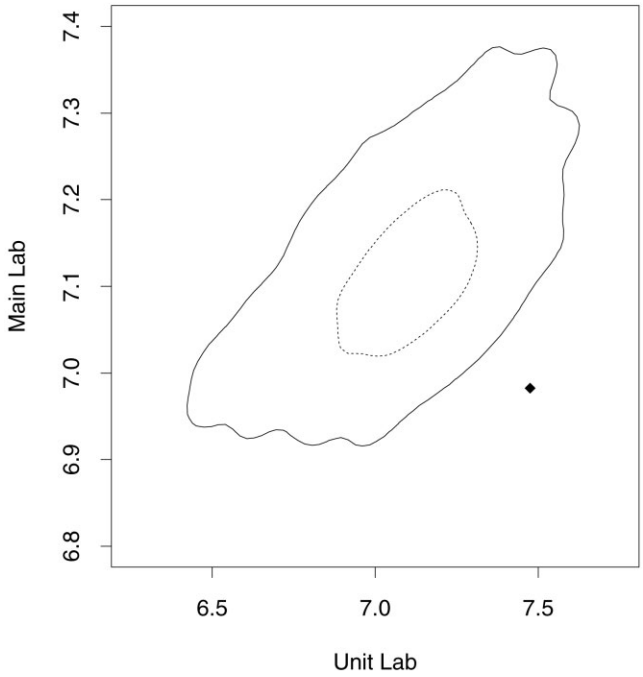

Figure 11. Equivalent control region of the density chart for the bivariate process of $\hat{\psi}_i = (\bar{X}_i, \bar{Y}_i)'$ for the lab measurement data. The area inside the solid curve is the control region. The dashed curve is the contour corresponding to the centerline

A contour plot of the bivariate kernel density estimate $\tilde{h}_{\hat{T}_\phi}(x)$ is presented in Figure 10. We set $\alpha = 0.01$ and used bivariate numerical integration and a bisection root finding routine to find $\hat{c}_{0.01} = 0.185\,435$ and $\hat{c}_{0.50} = 6.810\,448$. We computed the equivalent control region for the $\hat{\psi}_i = (\bar{X}_i, \bar{Y}_i)$. This region is plotted in Figure 11, along with the region corresponding to the centerline. The new subgroup observation is plotted in Figure 11, from which we can observe that this subgroup is outside of the control region, and the process is therefore out of control.

5. A SIMULATION STUDY

We first compare the performance of the usual $\bar{X}$-chart versus our proposed non-parametric location chart. In this study we simulate a set of 100 training observations from a specified distribution. The distributions studied are a subset of the 16 normal mixture densities presented by Marron and Wand[34]. These are the standard normal (#1), skewed unimodal (#2), highly skewed (#3), kurtotic unimodal (#4), separated bimodal (#7), and the asymmetric bimodal (#8) distributions. From the training sample we constructed the kernel density estimate $\tilde{h}_{\hat{t}}(x)$ using the two-stage bandwidth selection method and 2000 bootstrap resamples. We also constructed the traditional $\bar{X}$-chart using the estimate of variability based on the range. A sequence of new subgroups of size $(n = 5)$ were then simulated for each method from a process shifted by a specified amount δ until an out-of-control condition was encountered. This algorithm was repeated 1000 times and the average run length of the methods was estimated by the average number of additional subgroups needed until the chart was out of control. We used $\alpha = 0.01$ and $\delta = 0.00, 1.25, 1.50, 1.75$.

The $\bar{X}$ chart is a powerful technique due to its use of the central limit theorem. Studies by Burr[7] and Schilling and Nelson[8] have shown that the $\bar{X}$ chart is very robust to the normality assumption, even for subgroup sizes as small as $n = 5$. Hence, the usual $\bar{X}$ chart provides a good indication of near optimal performance for this problem (without parametrically modeling each distribution). It is also important to note that it is difficult to compare the average run lengths of methods that do not have the same actual false alarm rate. This is the same situation as comparing the power functions of two tests with different type I error probabilities.

Table I. Estimated average run lengths for the traditional $\bar{X}$ chart and the non-parametric location chart for a process shift of size δ. The specified false alarm probability is $\alpha = 0.01$

		δ				
Distribution	Method	0.00	0.25	0.50	0.75	1.00
Standard normal	Parametric	111.7	57.6	17.0	5.9	3.0
	Non-parametric	83.4	51.5	15.9	5.8	2.9
Skewed unimodal	Parametric	98.2	49.8	11.2	3.3	1.8
	Non-parametric	96.3	31.4	7.6	2.7	1.5
Highly skewed	Parametric	65.6	29.1	11.2	5.6	3.0
	Non-parametric	148.2	58.1	22.6	9.6	5.1
Kurtotic	Parametric	70.9	34.4	9.8	3.6	1.8
	Non-parametric	96.1	50.1	12.6	4.3	2.0
Separated bimodal	Parametric	113.0	72.6	32.0	14.3	7.4
	Non-parametric	88.7	59.0	26.6	12.0	6.5
Asymmetric bimodal	Parametric	119.5	71.3	20.9	6.9	3.5
	Non-parametric	95.1	45.1	13.8	5.5	2.9

Table II. Estimated average run lengths for the traditional $\bar{R}$ chart and the non-parametric variation chart for a process variance of δ^2 times that of the in-control process. The specified false-alarm probability is $\alpha = 0.01$

		δ					
Distribution	Method	1.00	1.05	1.10	1.15	1.20	1.25
Standard normal	Parametric	94.2	74.6	63.2	46.6	35.3	22.4
	Non-parametric	107.8	72.2	49.4	35.8	22.8	17.6
Skewed unimodal	Parametric	45.0	35.9	29.5	24.8	19.9	15.0
	Non-parametric	98.8	68.8	41.1	32.0	22.7	18.1
Highly skewed	Parametric	16.7	14.9	12.7	12.2	10.0	9.6
	Non-parametric	36.5	24.0	16.0	12.5	10.3	8.3
Kurtotic	Parametric	17.1	15.3	13.8	12.3	10.7	9.5
	Non-parametric	151.6	80.9	58.2	39.3	27.9	21.5
Separated bimodal	Parametric	71.0	80.7	89.6	112.4	113.6	130.0
	Non-parametric	75.4	52.5	26.6	14.4	9.1	5.3
Asymmetric bimodal	Parametric	173.1	151.3	122.3	90.0	72.7	46.3
	Non-parametric	96.1	61.5	39.1	33.9	18.6	14.0

The results of the study are presented in Table I. When $\delta = 0$ the process is in control and the average run length should be around 100 for a specified $\alpha = 0.01$. One can observe from Table I that our proposed non-parametric method maintains a reasonable average run length for the case when $\delta = 0$, expect for the case of the highly skewed distribution where the average run length is somewhat larger. The corresponding average run lengths for the remaining cases when $\delta > 0$ are also quite reasonable. Hence, in this case the non-parametric method provides reasonable performance when compared with a very powerful method.

In a second study we used the same setup as above, but instead studied the average run lengths of the range chart of our non-parametric method versus the standard range chart that assumes normality. Here we studied a multiplicative change in the process variance by δ^2 where $\delta = 1.00, 1.05, 1.10, 1.15, 1.20, 1.25$. The results of this study are presented in Table II. Again the behavior of the non-parametric method is quite reasonable. The parametric normal method, however, suffers from some problems. For the highly skewed and kurtotic cases when $\delta = 1.00$ the average run lengths are both below 20. This indicates that even when the process is in control, the standard range chart will indicate an out-of-control situation about once every 20 subgroups.

Another interesting effect with the range chart is the case of the separated bimodal distribution. Note that in this case the average run lengths of the normal chart actually become larger as δ increases. A similar effect is seen with the asymmetric bimodal distribution where the average run lengths stay quite large until δ becomes very large. This effect appears to occur because of an upward bias in the control limits. When δ is small, almost all of the out-of-control conditions are due to a subgroup range falling below the lower control limit. As δ increases, the ranges are more likely to be between the control limits and the average run length increases. It is only when δ becomes much larger that the subgroup ranges are large enough to be detected by the upper limit of control chart. Note that in each of these cases the behavior of the non-parametric chart remains reasonable.

6. CONCLUDING DISCUSSION

In this paper we have presented a very general methodology for constructing control charts. The advantages of the method are that it is applicable to many process characteristics and distributions. In multimodal situations the method produces control regions that are more reasonable when compared with traditional methods. Through several examples, we have shown in the parametric case that the method often reduces to standard control chart techniques. The method is also easily adapted to the non-parametric setting using smoothing techniques and the bootstrap. In our examples, the non-parametric method gave results that were consistent with, and sometimes better than, the standard techniques. In the simulations the method was often competitive with, if not superior to, the standard techniques. As such, we feel that this method has a great deal of potential in the area of statistical process control.

The non-parametric methods presented here can be applied to multivariate problems of small to moderate dimension. High-dimensional non-parametric estimation requires very large sample sizes that are often impractical for applied use. For more information on this problem, see Chapter 7 of Scott[17]. Of course, at the risk of losing some interpretability one may attempt to reduce the dimension of the problem using methods such as principal components before applying the proposed non-parametric method.

We feel that the non-parametric methods presented in this paper have the potential to satisfy all four of the guiding principals presented by Jackson[3]. Certainly the method produces a single answer as to whether the multivariate process is in control. Further, the method takes into account the relationships between the variables by modeling the sampling distribution of $\hat{\psi}_i$ directly, instead of through a parametric model. The remaining principals are the topic of current research. One would not expect the bootstrap method to exactly maintain the specified type I error rate for finite sample sizes. However, the in-depth research that has been done on the properties of bootstrap confidence intervals indicates that we can expect the type I error rate to be maintained at least in an asymptotic sense, with the error reasonably small for moderate training sample sizes[12]. Some indication of this type of behavior is observed in the simulation study where the type I error rate remained reasonable for many of the situations studied. However, further refinements of the method may be necessary to improve this behavior. In particular, corrections to the bootstrap method similar to those used for confidence intervals may be applicable here (see Efron[35] and Loh[36]). This is the topic of current research. Finally, as with the construction of multivariate parametric control charts, additional advances may be necessary to better understand and interpret an out-of-control condition with these charts. Another aspect of further research deals with the selection of the bandwidth matrix. While the methods we have chosen produce reasonable results, alternative bandwidth selection methods may improve the method. In particular, if a convex control region for $\hat{\psi}_i$ is desired, then it is clear from Figures 9 and 11 that our bandwidth may be too small. For further discussion of these difficulties, see p. 161 of Hall[12].

REFERENCES

1. Montgomery DC. *Introduction to Statistical Quality Control* (4th edn). Wiley: New York, 2002.
2. Alt FB. Multivariate quality control. *Encyclopedia of Statistical Sciences*, vol. 6, Kotz S, Johnson NL (eds.). Wiley: New York, 1985; 110–122.

3. Jackson EJ. Multivariate quality control. *Communications in Statistics—Theory and Methods* 1985; **14**:2657–2688.
4. Mason RL, Young JC. *Multivariate Statistical Process Control with Industrial Applications*. SIAM: Philadelphia, PA, 2002.
5. Randles RH, Wolfe DA. *Introduction to the Theory of Nonparametric Statistics*. Wiley: New York, 1999.
6. Alloway JA, Raghavachari M. Control chart based on the Hodges–Lehmann estimator. *Journal of Quality Technology* 1991; **23**:336–347.
7. Burr IW. The effect of nonnormality on constants for $\bar{X}$ and R charts. *Industrial Quality Control* 1967; **23**:563–569.
8. Schilling EG, Nelson PR. The effect of non-normality on the control limits of $\bar{X}$ charts. *Journal of Quality Technology* 1976; **8**:183–188.
9. Yourstone SA, Zimmer WJ. Nonnormality and the design of control charts for averages. *Decision Sciences* 1992; **23**:1099–1113.
10. Chou YM, Polansky AM, Mason RL. Transforming non-normal data to normality in statistical process control. *Journal of Quality Technology* 1998; **30**:133–141.
11. Nichols MD, Padgett WJ. A bootstrap control chart for Weibull percentiles. *Quality and Reliability Engineering International* 2005; (to appear).
12. Hall P. *The Bootstrap and Edgeworth Expansion*. Springer: New York, 1992.
13. Langenberg P, Iglewicz B. Trimmed mean $\bar{X}$ and R charts. *Journal of Quality Technology* 1986; **18**:152–161.
14. Bai DS, Choi IS. $\bar{X}$ and R charts for skewed populations. *Journal of Quality Technology* 1995; **27**:120–131.
15. Jones LA, Woodall WH. The performance of bootstrap control charts. *Journal of Quality Technology* 1998; **30**:362–375.
16. Bowman AW, Azzalini A. *Applied Smoothing Techniques for Data Analysis*. Clarendon Press: Oxford, 1997.
17. Scott DW. *Multivariate Density Estimation*. Wiley: New York, 1992.
18. Simonoff JS. *Smoothing Methods in Statistics*. Springer: New York, 1996.
19. Wand MP, Jones MC. *Kernel Smoothing*. Chapman and Hall: London, 1995.
20. Efron B. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics* 1979; **7**:1–26.
21. Seppala T, Moskowitz H, Plante R, Tang J. Statistical process control via the subgroup bootstrap. *Journal of Quality Technology* 1995; **27**:139–153.
22. Bajgier SM. The use of bootstrapping to construct limits for control charts. *Proceedings of the Decision Science Institute*, San Diego, CA, 1992. Decision Science Institute: Atlanta, GA, 1992; 1611–1613.
23. Liu RY, Tang J. Control charts for dependent and independent measurements based on the bootstrap. *Journal of the American Statistical Association* 1996; **91**:1694–1700.
24. Polansky AM. Stabilizing bootstrap-t confidence intervals for small samples. *Canadian Journal of Statistics* 2000; **59**:821–838.
25. Sheather SJ, Jones MC. A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society B* 1991; **53**:683–690.
26. Silverman BW. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall: London, 1986.
27. Conte SD, de Boor C. *Elementary Numerical Analysis*. McGraw-Hill: New York, 1980.
28. Deheuvels P. Estimation nonparamétrique de la densité par histogrammes généralisés (II). *Publications de l'Institut Statistique de l'Université de Paris* 1977; **22**:1–23.
29. Wand MP, Jones MC. Multivariate plug-in bandwidth selection. *Computational Statistics* 1994; **9**:97–117.
30. Wand MP, Jones MC. Comparison of smoothing parameterizations in bivariate kernel density estimation. *Journal of the American Statistical Association* 1993; **88**:520–528.
31. Liu RY. Control charts for multivariate processes. *Journal of the American Statistical Association* 1995; **90**:1380–1387.
32. Liu RY. On a notion of data depth based on random simplicies. *The Annals of Statistics* 1990; **18**:405–414.
33. Tracy ND, Young JC, Mason RL. A bivariate control chart for paired measurements. *Journal of Quality Technology* 1995; **27**:370–376.
34. Marron JS, Wand MP. Exact mean integrated squared error. *The Annals of Statistics* 1992; **20**:712–736.
35. Efron B. Better bootstrap confidence intervals. *Journal of the American Statistical Association* 1987; **82**:171–200.
36. Loh WY. Calibrating confidence coefficients. *Journal of the American Statistical Association* 1987; **82**:155–162.

Author's biography

Alan M. Polansky earned a PhD in Statistical Science from Southern Methodist University in 1995. He also has an MS and BS in Mathematics from the University of Texas at San Antonio. He is currently an Associate Professor in the Division of Statistics at Northern Illinois University. His research in the area of quality has focused on the development of non-parametric approaches. This includes using modern non-parametric

techniques such as the bootstrap and kernel smoothing to solve problems in process capability and process monitoring. He is a member of the American Statistical Association, the Statistical Society of Canada and the American Society for Quality. He is also a founding member and executive board member of the Section on Nonparametric Statistics of the American Statistical Association.