Mahmoud, M. A., Parker, P. A., Woodall, W. H., and Hawkins, D. M. (2007), "A Change Point Method for Linear Profile Data," *Quality and Reliability Engineering International*, 23, 247–268. [265]
Rice, J. A., and Wu, C. O. (2001), "Nonparametric Mixed Effects Models for Unequally Sampled Noisy Curves," *Biometrics*, 57, 253–259. [266]
Seifert, B., Brockmann, M., Engel, J., and Gasser, T. (1994), "Fast Algorithms for Nonparametric Estimation," *Journal of Computational and Graphical Statistics*, 3, 192–213. [270]
Shi, M., Weiss, R. E., and Taylor, J. M. G. (1996), "An Analysis of Paediatric CD4 Counts for Acquired Immune Deficiency Syndrome Using Flexible Random Curves," *Applied Statistics*, 45, 151–163. [266]
Walker, E., and Wright, S. P. (2002), "Comparing Curves Using Additive Models," *Journal of Quality Technology*, 34, 118–129. [266]
Wang, Y. (1998), "Smoothing Spline Models With Correlated Random Errors," *Journal of the American Statistical Association*, 93, 341–348. [267]
Williams, J. D., Birch, J. B., Woodall, W. H., and Ferry, N. M. (2007), "Statistical Monitoring of Heteroscedastic Dose-Response Profiles From High-Throughput Screening," *Journal of Agricultural, Biological, and Environmental Statistics*, 12, 216–235. [265]
Williams, J. D., Woodall, W. H., and Birch, J. B. (2007), "Statistical Monitoring of Nonlinear Product and Process Quality Profiles," *Quality and Reliability Engineering International*, 23, 925–941. [265,266,273]
Woodall, W. H. (2000), "Controversies and Contradictions in Statistical Process Control," *Journal of Quality Technology*, 32, 341–350. [265]
Woodall, W. H., Spitzner, D. J., Montgomery, D. C., and Gupta, S. (2004), "Using Control Charts to Monitor Process and Product Quality Profiles," *Journal of Quality Technology*, 36, 309–320. [265]
Wu, H., and Zhang, J. (2002), "Local Polynomial Mixed-Effects Models for Longitudinal Data," *Journal of the American Statistical Association*, 97, 883–897. [266-268,270,271,275]
Zhang, D., Lin, X., Raz, J., and Sowers, M. (1998), "Semiparametric Stochastic Mixed Models for Longitudinal Data," *Journal of the American Statistical Association*, 93, 710–719. [267]
Zou, C., Tsung, F., and Wang, Z. (2007), "Monitoring General Linear Profiles Using Multivariate EWMA Schemes," *Technometrics*, 49, 395–408. [265, 266,273]
——— (2008), "Monitoring Profiles Based on Nonparametric Regression Methods," *Technometrics*, 50, 512–526. [265,271,273,274]
Zou, C., Zhang, Y., and Wang, Z. (2006), "Control Chart Based on Change-Point Model for Monitoring Linear Profiles," *IIE Transactions*, 38, 1093–1103. [265]

Comment

Daniel W. Apley

Department of Industrial Engineering
and Management Sciences
Northwestern University
Evanston, IL 60208-3119
(*apley@northwestern.edu*)

I congratulate the authors on a welcome addition to the profile monitoring literature, particularly regarding how to account for within-profile correlation. I cannot imagine a set of real profile data that does not have within-profile correlation, at least not for $\{x_{ij}: j = 1, 2, \ldots, n_i\}$ that are densely enough spaced to be consistent with the spirit of profile data. Although within-profile correlation may have little adverse effect when fitting linear or simple parametric nonlinear models, as noted by Jensen, Birch, and Woodall (2008), it is inherently more insidious when fitting nonparametric models that are based on local smoothing. Profile variation due to a random component that is correlated within-profile looks deceptively like local, nonrandom changes in the mean if the model assumes no within-profile correlation. It is, therefore, likely to cause excessive false alarms. Although many authors noted the pervasiveness of within-profile correlation, surprisingly few developed algorithms that take this into account. Notable exceptions are the mixed model approach of Jensen, Birch, and Woodall (2008) and Jensen and Birch (2009) and the spatial autoregressive approach of Colosimo, Semeraro, and Pacella (2008). In light of this, I think the present work will be a welcome addition for practitioners who wish to monitor nonlinear profiles that are too irregularly shaped to be modeled parametrically.

I can find very little to criticize, but I would like to direct further scrutiny to three issues. The first two regard choosing the control limits to avoid excessive false alarms. These are hardly criticisms because I believe the authors' resampling procedure (with suitable modification) presents a nice solution to this dilemma. The third is a broader issue that regards the nature of the assignable causes that are typically assumed in the profile monitoring literature.

BEWARE OF ASYMPTOTIC RESULTS

In general, asymptotic results are often very useful in statistics. Take the central limit theorem, for example. One reason it is so useful is that the conditions under which we can approximate nonasymptotic reality using the asymptotic results of the central limit theorem are often met in practice: Usually only moderate sample sizes are needed to approximate the distribution of the average of a random sample as normal, at least for the level of accuracy required in many applications.

Regarding the asymptotic results of Theorem 1, on the other hand, I have doubts that the conditions required for their approximate validity are satisfied for typical profile monitoring applications. One conclusion that the authors draw following Theorem 1 is that the asymptotic distribution of $T_{t,h,\lambda}$ is independent of the "nuisance" parameter $\gamma(x, x')$ when the condition of bounded n_ih is met (which is always the case in practice). The implication seems to be that one may choose the

TECHNOMETRICS, AUGUST 2010, VOL. 52, NO. 3
DOI 10.1198/TECH.2010.09151

control limits independent of $\gamma(x, x')$ and achieve a false alarm probability that is insensitive to uncertainty in this nuisance parameter (just prior to Theorem 1, the authors state that it can "*shed some light on practical design of the chart*"). But $\gamma(x, x')$ is the covariance function of $f_i(x)$, which represents the component of profile variation that accounts for within-profile correlation. It seems almost obvious that within-profile correlation can strongly affect the false alarm probability, for the simple reason that I state in the first paragraph of this discussion. Tighter arguments follow by noting that Equation (9) implies that the estimated profile mean $\widehat{g}_{t,h,\lambda}(s)$ is a locally weighted average of the observed profile values y_{ij} within some kernel neighborhood of s. Certainly, the variance of a weighted average depends strongly on the correlation between the variables that are being averaged. Because the weights are likely to be mostly positive, the variance of $\widehat{g}_{t,h,\lambda}(s)$ will be much larger if $\gamma(x, x')$ represents high positive correlation between $f(x)$ and $f(x')$ for x and x' both within the neighborhood of s than if $\gamma(x, x')$ represents no within-profile correlation. A larger variance of $\widehat{g}_{t,h,\lambda}(s)$ will, in turn, increase the mean of $T_{t,h,\lambda}$, perhaps substantially. Indeed, if the distribution of $T_{t,h,\lambda}$ truly were approximately independent of $\gamma(x, x')$, then the primary motivation for this work (handling within-profile correlation) will disappear.

So what explains this seeming contradiction that the in-control distribution of $T_{t,h,\lambda}$ depends strongly on within-profile correlation in most practical scenarios but has no dependence whatsoever asymptotically? The explanation is that, generally speaking, it can be very difficult to take asymptotic results for complex models involving a long list of convoluted conditions and render them down to their practical implications. In the present case, a careful inspection of the list of regularity conditions in Appendix A reveals some unrealistic ones that explain the contradiction at hand and account for other questionable conclusions that one might be tempted to draw from Theorem 1. For example, part of Condition C8 is that the kernel bandwidth $h \to 0$. In other words, the neighborhoods over which the kernel-weighted smoothing for $\widehat{g}_{t,h,\lambda}(s)$ takes place must shrink down to infinitesimally small neighborhoods. This condition is clearly at the heart of the invariance of the asymptotic distribution of $T_{t,h,\lambda}$ to $\gamma(x, x')$. For infinitesimally small neighborhoods, more of the y_{ij} in Equation (9) come from different i [i.e., more averaging across *different* profiles—see the next paragraph for why the exponentially weighted moving average (EWMA) time window is infinitely long in Theorem 1] and fewer from different j (i.e., less averaging *within* profiles). Hence, the asymptotic conditions of Theorem 1 do not even correspond to using local smoothing. They correspond to simply averaging y_{ij} over the time index i at a fixed spatial index j when estimating $g(\cdot)$. This is why the asymptotic results of Theorem 1 have the distribution of $T_{t,h,\lambda}$ independent of $\gamma(x, x')$.

Of course, this has little practical relevance because one will never choose infinitesimally small spatial neighborhoods and infinitely large time windows. It will be interesting to see how $\gamma(x, x')$ affects the distribution of $T_{t,h,\lambda}$ for typical values of h and λ commensurate with those recommended in Section 2.5.

A related concern is that Condition C8-II implies that both $\lambda \to 0$ and $t \to \infty$, if n_i is bounded (which is always the case in practice, no matter how large it is). This is entirely unrealistic because it implies that the EWMA is using an infinitely long time window. In practice, to remain responsive to sudden large shifts, moderate length time windows (e.g., corresponding to $0.05 \le \lambda \le 0.2$) are usually chosen. As I discuss in the preceding paragraphs, the infinitely long time window (coupled with the infinitesimally small spatial neighborhood) explains why the asymptotic distribution of $T_{t,h,\lambda}$ in Theorem 1 is invariant to within-profile correlation. It also explains the asymptotic normality of $T_{t,h,\lambda}$ in Theorem 1. I suspect that for typical choice of λ, the distribution of $T_{t,h,\lambda}$ is quite positively skewed (as are distributions of many quadratic forms).

My main point here is that readers will be well advised not to assign too much significance to Theorem 1 and certainly not to select control limits based on the asymptotic distribution of $T_{t,h,\lambda}$ that Theorem 1 implies. Fortunately, readers do not have to because the authors included a very attractive resampling approach for calculating appropriate control limits. My next comment is related to this.

TEMPORAL AUTOCORRELATION AND FALSE ALARMS

The availability of spatially dense (i.e., large n_i) profile data is often the result of sophisticated, automated measurement technology. But this also often results in temporally dense profile data, for which a large number of profiles are collected over a relatively short period of time. Temporally dense profile data are likely to have temporal autocorrelation, in addition to the spatial autocorrelation represented by $\gamma(x, x')$.

It is well known that (positive) temporal autocorrelation causes a dramatic increase in the false alarm rate of many univariate control charts for detecting mean shifts. This is especially true for charts that have longer memory, such as a cumulative sum (CUSUM) with a small reference value or an EWMA with small λ (Apley and Lee 2003, 2008). Because of the manner in which the EWMA is involved in Equation (9), positive temporal autocorrelation will tend to increase the variance of $\widehat{g}_{t,h,\lambda}(s)$, thereby increasing the mean of $T_{t,h,\lambda}$. This follows from reasoning similar to that discussed in the preceding section in the context of spatial autocorrelation. An increase in the mean of $T_{t,h,\lambda}$ will, in turn, increase the false alarm rate, perhaps substantially.

Consequently, for many practical profile monitoring applications, I suspect that the control limits will have to be altered (widened) to account for the autocorrelation. We have simple time-series models for effectively representing temporal autocorrelation in the univariate control charting case, and these models offer means of appropriately altering the control limits. However, it seems doubtful that the model in Equation (1) can be augmented in a tractable yet realistic manner to represent temporal autocorrelation in the profile monitoring case. The authors' resampling approach in its present form would not result in properly widened control limits when autocorrelation is present, because the completely randomized profile resampling destroys any temporal structure in the data.

Fortunately, bootstrap resampling procedures can be modified to take into account temporal autocorrelation. Two main approaches for accomplishing this are the Markov bootstrap procedure of Paparoditis and Politis (2002) and the block bootsrap procedure of Künsch (1989). In the Markov bootstrap procedure, one assumes the autocorrelation can be described in terms of a vector of random variables that follows a vector Markov process with unknown transition probability function. Essentially, one uses a form of kernel density estimation to fit the transition probability function, which one then uses to govern the sequential resampling in a manner that attempts to preserve the autocorrelation. In the block bootstrap procedure, which entirely avoids the need to model the autocorrelation, one will resample the profiles in blocks, instead of one-at-a-time. Each block will consist of a temporally contiguous set of profiles, the first of which is chosen randomly. Shan and Apley (2008) provided a more detailed description of the two procedures for a related problem.

The main drawback of the Markov bootstrapping procedure is that one must identify a Markov vector of variables that can represent the autocorrelation. The vector must have a dimension low enough to allow kernel density estimation of the transition probability function (e.g., one- or two-dimensional), and the transition probability function must be reasonably close to Markov. It is doubtful that the nonparametric structure of Equation (1) will yield a suitable low-dimensional Markov vector that can realistically account for the temporal autocorrelation in typical profile data. The main drawback of the block bootstrap procedure is that the total number of profiles [m in Equation (1)] must be relatively large. The length of each individual block should be only a fraction of m (e.g., less than 10%), while each individual block should be long enough to allow the dynamics due to the autocorrelation within each block to dominate the discontinuities between blocks. The value $m \geq 500$ recommended in Section 2.5 might be sufficient for the block bootstrap procedure in most cases.

In light of this, the block bootstrap procedure will most likely be more appropriate than the Markov bootstrap procedure for setting the control limits when monitoring temporally autocorrelated profiles. For the authors' real-data application, it will be interesting to see if temporal autocorrelation were present. This can be easily assessed graphically by choosing a few values of s, and then constructing time series charts of $\widehat{g}(s) + \widehat{f}_i(s)$ versus i for $i = 1, 2, \ldots, m$. If temporal autocorrelation appears substantial, it will also be interesting to see if the block bootstrap procedure results in wider control limits than the regular bootstrapping procedure.

FOR WHAT SHOULD WE BE LOOKING?

The objective in this article, as well as in most of the profile monitoring literature, is to detect a change in the profile mean function $g(\cdot)$. To borrow statistical process control (SPC) terminology, in the model $y_{ij} = g(x_{ij}) + f_i(x_{ij}) + \varepsilon_{ij}$, the authors view changes in $g(\cdot)$ as assignable causes of variation and $f_i(x_{ij}) + \varepsilon_{ij}$ as common causes of variation. For many applications, I imagine changes in $g(\cdot)$ are important (perhaps the most important) indicators of assignable causes, and more generally, of problems with the process that should be detected and corrected. On the other hand, there are many applications in which assignable causes are manifested as something other than a change in the mean.

Within the realm of the model in Equation (1), the next obvious characteristic to monitor is a change in the covariance $\gamma(x, x')$. Jin and Shi (1999) considered stamping tonnage profile monitoring (each profile is the press tonnage signature for one cycle, corresponding to one stamped part) and provide an illuminating discussion of a number of typical assignable causes and the effects they have on the profiles. Although some are manifested as changes in the profile mean, many others will be more reasonably represented as increased variation in $f_i(x)$ at one or more x, which corresponds to an increase in $\gamma(x, x)$. The authors of the present article mention detecting changes in $\gamma(x, x')$ in their conclusions, but leave it as future work due to its nontrivial nature. However, detecting certain changes in $\gamma(x, x')$ may, in fact, require only a trivial modification of their algorithm. Because of the quadratic nature of $T_{t,h,\lambda}$, I suspect that their algorithm will be reasonably effective at detecting variance changes [i.e., changes in $\gamma(x, x)$] if we forego the EWMA by using $\lambda = 1$ when estimating $\widehat{g}_{t,h,\lambda}(s)$. In addition to using $\lambda = 1$ when estimating $\widehat{g}_{t,h,\lambda}(s)$, we might impart memory in a different way by incorporating exponential weighting directly into the equation for $T_{t,h,\lambda}$. Specifically, using the authors' notation, we might consider

$$\sum_{i=0}^{t-1} (1 - \rho)^i T_{t-i,h,1},$$

as a control chart statistic, where $0 < \rho \leq 1$ denotes another EWMA parameter, and the third subscript on $T_{t,h,1}$ indicates that we use $\lambda = 1$ when estimating $\widehat{g}_{t,h,\lambda}(s)$. We can view this as analogous to an exponentially weighted moving variance.

I doubt that monitoring for generic changes in the covariance structure $\gamma(x, x')$ beyond the variance $\gamma(x, x)$ will be fruitful. In analogy with T^2 control charts for high-dimensional multivariate data (viewing each profile as a vector), for real datasets I envision such a chart plagued by alarms caused by innocuous changes in the correlation between $f_i(x)$ and $f_i(x')$ with no appreciable change in their variances. Even though these represent legitimate changes in the within-profile covariance, I believe most practitioners will prefer to view them as nuisance alarms.

Regarding looking outside the realm of Equation (1) for assignable causes, from one perspective it is unnecessary, because the nonparametric nature of Equation (1) makes it almost completely generic. Indeed, any set of profiles can be represented as a mean function plus a deviation from the mean that is zero-mean (by definition) with some covariance function. Likewise, almost any change in the profiles that we might care about will result in a change in the mean and/or covariance. However, using an algorithm based on the generic model in Equation (1) may be far from the most effective way of detecting such changes. Referring again to the stamping example of Jin and Shi (1999), some of their identified assignable causes did result in changes in the profile mean, but they were very specific changes of known form (e.g., a reduction in the peak tonnage) that can be premodeled. This is analogous to knowing in advance the direction of a change in the mean vector in multivariate control charting. Taking such information into account will greatly enhance the power to detect such changes.

Other assignable causes in Jin and Shi (1999) resulted in oscillation of the tonnage signature that amounted to highly structured changes in $\gamma(x, x')$. Again, an algorithm that incorporates knowledge of the specific structure of the change can result in much more powerful detection. With limited, incomplete knowledge of the structure of the change, monitoring coefficients of a Fourier or wavelet representation of the profiles can sometimes be useful (see Chicken, Pignatiello, and Simpson 2009, and the references therein).

In general, premodeling potential assignable causes and their effects on the profiles may be quite difficult for many applications, requiring advanced engineering knowledge and resources. It will be useful to have better "Phase I" exploratory data analysis tools for discovering and empirically modeling the effects of typical assignable causes based on large historical sets of profiles, over which various assignable causes occurred. It will also be useful to have an approach that looks specifically for a small set of patterns that might be easily premodeled, while simultaneously monitoring for more general profile changes via a $T_{t,h,\lambda}$-like statistic. Apley and Lee (2010) developed a related approach for multivariate process data, but this will be difficult to extend to profile data.

I will close by thanking the authors for a thought-provoking article and a useful approach that I hope will find its way into SPC practitioners' toolboxes. I would also like to thank the editor, David Steinberg, for recognizing the merit of their work and inviting these discussions.

ACKNOWLEDGMENT

This work was partially supported by the National Science Foundation under grant CMMI-0826081.

ADDITIONAL REFERENCES

Apley, D. W., and Lee, H. C. (2003), "Design of Exponentially Weighted Moving Average Control Charts for Autocorrelated Processes With Model Uncertainty," *Technometrics*, 45 (3), 187–198. [278]

——— (2008), "Robustness Comparison of Exponentially Weighted Moving Average Charts on Autocorrelated Data and on Residuals," *Journal of Quality Technology*, 40 (4), 428–447. [278]

Apley, D. W., and Lee, H. Y. (2010), "Simultaneous Identification of Premodeled and Unmodeled Variation Patterns," *Journal of Quality Technology*, 42, 36–51. [280]

Chicken, E., Pignatiello, J. J., and Simpson, J. R. (2009), "Statistical Process Monitoring of Nonlinear Profiles Using Wavelets," *Journal of Quality Technology*, 41 (2), 198–212. [280]

Colosimo, B. M., Semeraro, Q., and Pacella, M. (2008), "Statistical Process Control for Geometric Specifications: On the Monitoring of Roundness Profiles," *Journal of Quality Technology*, 40 (1), 1–18. [277]

Jin, J., and Shi, J. (1999), "Feature-Preserving Data Compression of Stamping Tonnage Information Using Wavelets," *Technometrics*, 41 (4), 327–339. [279,280]

Künsch, H. R. (1989), "The Jackknife and the Bootstrap for General Stationary Observations," *The Annals of Statistics*, 17 (3), 1217–1241. [279]

Paparoditis, E., and Politis, D. N. (2002), "The Local Bootstrap for Markov Processes," *Journal of Statistical Planning and Inference*, 108, 301–328. [278]

Shan, X., and Apley, D. W. (2008), "Blind Identification of Manufacturing Variation Patterns by Combining Source Separation Criteria," *Technometrics*, 50 (3), 332–343. [279]

Comment

Hugh A. Chipman

Department of Mathematics and Statistics
Acadia University
Wolfville, NS B4P 2R6
Canada
(*hugh.chipman@acadiau.ca*)

R. Jock MacKay and Stefan H. Steiner

Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, ON N2L 3G1
Canada

We thank the editor for the opportunity to be discussants and congratulate the authors on a stimulating article.

Profile monitoring is an area of growing interest and importance. The authors develop a methodology that meets many of the needs of practitioners. They propose a flexible model based on a solid statistical foundation. Nonparametric local regression methods and random effects form the core of their approach. The random effects provide a convenient way of modeling covariance between responses observed at different points along the curve, a common feature of functional data. The procedure is quick in Phase II and appears to readily adapt to a variety of profile shapes.

To organize our discussion, we attempt to outline a list of desirable attributes and questions we can ask of a profile monitoring methodology. After describing each, we examine Qiu, Zou, and Wang in the context of that attribute or those questions. Before presenting our list, we briefly discuss a motivating example.

Example. To help fix ideas and provide a broader basis for discussing desirable attributes, we briefly describe a profile monitoring problem familiar to us. Mosesova (2007) provides additional details. The data arise from a manufacturing process in which a ram force-fits a steel valve seat into an aluminum cylinder head. Every insertion yields a force–time profile, three of which are displayed in Figure 1. In this particular process, a feedback controller adjusts the force in an attempt to maintain constant ram velocity during insertion. After an initial rise in

TECHNOMETRICS, AUGUST 2010, VOL. 52, NO. 3
DOI 10.1198/TECH.2010.09180