A Management Framework for Process Control and Capability In The Presence Of Autocorrelation

Daniel J. Zalewski, Ph.D., Office of the Secretary of Defense
Edward F. Mykytka, Ph.D., University of Dayton

Abstract

The most commonly used tools for managing manufacturing quality and productivity – control charts – are designed to maintain a process in a state of statistical control, which implies that product variation is at a minimum and due solely to a constant system of chance causes. This, in turn, ensures that the product is as consistent as possible and, in a well-designed system, is capable of meeting or exceeding specifications.

Standard control charts, however, are founded on the intuitive assumption that successive measurements taken on an in-control process behave as a sequence of independent and identically distributed random variables. While this may be true for traditional discrete-part manufacturing systems wherein measurements are taken infrequently, it is likely not true for continuous-flow systems or for modern discrete-part environments with fast production rates, highly interconnected processes, and automatic measurement systems. Measurements taken within these latter environments are often highly autocorrelated and, if this dependence is ignored, can render standard control charts and process capability analyses invalid. Even when accounted for, the autocorrelation between measurements can lead to incomplete or misleading assessments of process capability and control.

To overcome these shortcomings, we extend the existing definitions of common and special cause variation to explicitly, and more intuitively, account for the variation induced by autocorrelation and propose a new management framework for monitoring process capability rather than statistical control. We differentiate between short-term and long-term capability and briefly demonstrate an implementation of this framework.

Keywords: Autocorrelation, Control Charts, Process Capability, Statistical Process Control, Time Series Models

Introduction

The objective of this paper is to introduce a new framework for monitoring and ensuring the quality of the output from modern manufacturing processes. We first review the tenets of statistical process control and process capability analysis and argue that the concepts of capability and control—which traditionally and intentionally have been regarded as separate and distinct—become intimately entwined when the process output is correlated over time. Our goal is to initiate discussion among leading managers of technology as to the merits of this approach.

Process Variation and Control

The quality of the output from a manufacturing process is typically managed by, first, designing a process that is capable of creating products that meet customer requirements and, then, maintaining this process in a "capable" state. These tasks, however, are complicated by the fact that all processes vary. The key to success is understanding and controlling this variation.

The tools customarily used to manage variation are founded on the classification of the causes of variation into two types. The first type, called assignable or special causes, are those factors that contribute most greatly to observed variation and which can be detected, identified, and eliminated. These are usually relatively few in number and typically result from human error, defective inputs, or improperly adjusted processes. The second type of causes are called chance or common causes and are those factors, generally numerous and individually of small importance, that are not feasible to detect, identify, or control. These are viewed as the sources of the natural or inherent variability within a system. [These definitions (and the others that follow) are consistent with traditional use in statistical quality control and adhere to the lexicon established by the American Society for Quality (1996)].

A process that operates free of special cause variation is said to be "in a state of statistical control" or, more simply, "in control." Shewhart (1939), who is credited with founding the field of statistical process control, describes the situation as follows:

A phenomenon will be said to be controlled when . . . we can predict, at least within limits, how the phenomenon may be expected to vary in the future. Here is it understood that prediction within limits means that we can state, at least approximately, the probability that the observed phenomenon will fall within the given limits.

To describe this state of affairs mathematically, let x_1, x_2, x_3, . . . denote the measured values of a relevant quality characteristic from a sequence of items produced by the process. Then, when only common cause variation is present, these observations are commonly assumed to arise from a model of the form

$$x_i = \mu + \varepsilon_i,$$

where the "process mean" μ is the average measurement that this process will produce when it is in control and ε_i is a random variable that represents the amount by which the i^{th} measurement will differ from this mean owing to the effects of common causes. Typically, $\varepsilon_1, \varepsilon_2, \varepsilon_3$, . . . are regarded as a sequence of independent and identically distributed random variables with mean zero and standard deviation σ and the parameters μ and σ are estimated from observation of the process when it is presumed to be in control.

When the process is in control, it follows that x_1, x_2, x_3, . . . are independent and identically distributed random variables with mean μ and standard deviation σ. In turn, we can make a prediction about the future behavior of this sequence by noting that, as long as the process is in control, the vast majority of future measurements will lie within the natural tolerance limits defined by $\mu - 3\sigma$ and $\mu + 3\sigma$. In fact, when we additionally assume that the measurements are normally distributed, it turns out that 99.73% of all future measurements can be expected to take on values between these limits as long as the process remains in control.

Conversely, a process that is operating in the presence of assignable causes is said to be out of control since these assignable causes can and should be eliminated. To do so, however, we must be able to detect their presence. Fortunately, assignable causes typically manifest themselves as shifts in the process mean and/or standard deviation away from their in-control values, making it more likely that measurements will lie outside the natural tolerance limits. Thus, whenever a measurement falls outside the $\mu \pm 3\sigma$ limits, now regarded as 3-sigma control limits, it is evidence that an assignable cause has occurred. Good practice then requires the process to be stopped and a search for an assignable cause to be initiated.

A control chart is simply a plot of the measurements x_1, x_2, x_3, . . . over time relative to their control limits and provides an immediate visual check on whether or not the process is in control. Such a plot also enables one to detect non-random patterns in the measurements (other than points plotting outside the control limits) that may be indicative of the presence of an assignable cause. For example, a sequence of at least eight points plotting above the process mean suggests the possible presence of an assignable cause, since there is a very small probability (about 0.4%) of this occurring due to chance alone when the measurements are indeed independent and identically distributed random variables as we assume when in control.

Figure 1 displays a control chart with 3-sigma limits for a process within which the first 50 (simulated) measurements were produced when the process was in control (with mean $\mu = 0$ and standard deviation $\sigma = 1$) and the second 50 measurements were produced after an assignable cause shifted the mean to $\mu = 1$. Note that, although no points plot outside the control limits, the run of 19 consecutive measurements above the in-control mean (starting with measurement 56) emphatically suggests that the process is out of control.

Process Capability

Next suppose that customer requirements can be summarized by stipulating a target (or

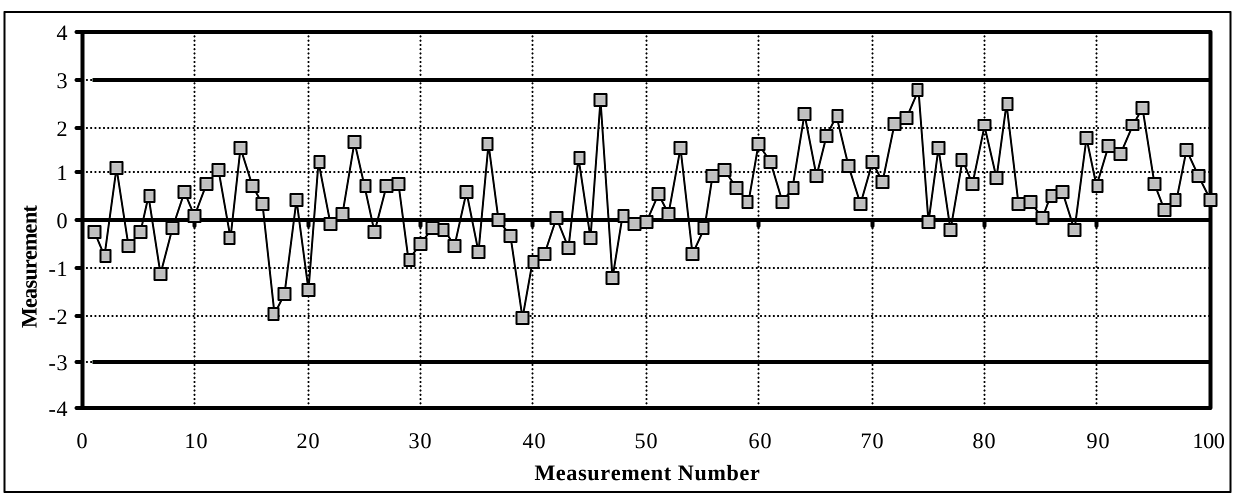

Figure 1: Uncorrelated Measurements; Shift Begins after Measurement Number 50

ideal) value, T, and lower and upper specification limits, LSL and USL, for the measured quality characteristic. In this context, an ideal process would be one whose process mean is the same as the target and whose measurements do not vary from that value. More realistically, a process that is capable of consistently meeting customer requirements is one whose natural tolerance limits lie within the specification limits so that, when the process is in control, the vast majority of the items produced will meet requirements. As a result, when process control methods are applied to capable processes, they not only ensure that the process remains in control, they also ensure that it remains capable of meeting customer requirements.

The design of a capable process requires the use of time-honored engineering and statistical experimental design methodologies and a means of measuring and assessing its ability to produce output that consistently meets customer requirements. Typically, the latter is accomplished by computing one of a number of common process capability indices. Two extensively used measures of process capability are given by

$$C \;= \frac{USL - LSL}{6\sigma} \qquad \text{and} \qquad C_{pk} = \mathrm{Min}\left\{\frac{USL - \mu}{3\sigma}, \frac{\mu - LSL}{3\sigma}\right\}.$$

Each of these indices relate the specified requirements limits to the inherent variability of the process (as measured by the width, 6σ, or half-width, 3σ, of the natural tolerance limits). Values of at least 1.33 or 1.5 are often recommended for these indices [Montgomery (1996)].

Controlling Autocorrelated Processes

The preceding discussions of capability and control are founded on the assumption that successive measurements from an in-control process behave as independent and identically distributed random variables. While this may be true for traditional discrete-part manufacturing systems wherein measurements are taken infrequently, it is likely <u>not</u> true for continuous-flow systems or for modern discrete-part environments with fast production rates, highly interconnected processes, and automatic measurement systems. Often in these environments, a measurement that is low relative to its mean tends to be followed by another relatively low measurement, while a relatively high measurement tends to be followed by another high one. In this case, the measurements are said to be positively autocorrelated. Unfortunately, as Montgomery (1996) writes, “Control charts do not work well if the quality characteristic exhibits even low levels of correlation over time.” Specifically, standard charts will have a higher rate of false alarms (points plotting outside of the control limits when the process is in control) and can take longer to detect an assignable cause when the data are correlated [Lu and Reynolds (1999)].

A common method used by many authors to account for the correlation between successive measurements taken on an in-control process is to assume that those measurements arise from a time series model of the form

$$x_i = \xi + \varphi_1 x_{i-1} + \varphi_2 x_{i-2} + \cdots + \varphi_p x_{i-p} + \varepsilon_i - \theta_1\varepsilon_{i-1} - \theta_2\varepsilon_{i-2} - \cdots - \theta_q\varepsilon_{i-q},$$

where, as before, $\varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots$ are a sequence of independent and identically distributed random variables with mean zero and standard deviation σ and the model parameters ξ, $\varphi_1, \varphi_2, \ldots, \varphi_p$, $\theta_1, \theta_2, \ldots, \theta_q$ are estimated from observation of the process when it is presumed to be in control.

This is commonly referred to as an "autoregressive-moving-average process of order p and q," denoted ARMA(p,q) [see Box and Jenkins (1976)].

Although complex in appearance, this model is very useful for modeling autocorrelated measurements and forms the basis for a number of process control methods. [See, for example, Alwan and Roberts (1989), Berthoux, Hunter and Pallsen (1976), Yourstone and Montgomery (1989), and Atienza, Tang, and Ang (1998).] These methods generally require one to use a time series model to forecast the next measurement and determine whether or not the process is in control by using a control chart to monitor the differences between the forecasts and the measurements actually observed.

Our objective here is to neither review nor critique these methods but to explore the implications of using a time series to model the behavior of an in-control process. Figure 2, for example, depicts 100 measurements from a chemical process, which Montgomery and Mastrangelo (1991) show are well fit by an ARMA(1,0) model with $\xi = 13.04$ and $\varphi_1 = 0.847$. Note that this series shows a clear tendency to meander. This pattern is typical of data from real-world autocorrelated processes and, we posit, would lead many quality managers—who have formed their intuition and expertise working with <u>un</u>correlated processes—to conclude that the process is out of control.

Indeed, the dashed lines in Figure 2, as originally displayed by Montgomery and Mastrangelo (1991), show 3-sigma control limits for an individuals chart within which the process standard deviation has been estimated from moving ranges. This chart, which would be appropriate if the measurements were uncorrelated, has many points plotting outside the control limits and, thus, clearly suggests that the process is out of control. In addition, Faltin and Woodall (1991) state that the "autocorrelation itself may indicate the presence of sources of variability which should be removed, rather than modeled, if possible."

On the other hand, Montgomery and Mastrangelo also show the process to be <u>in</u> control if one uses a time series model (or an approximation thereof) to account for the autocorrelation in the time series. This is reinforced by the wider control limits given by solid lines in Figure 2,

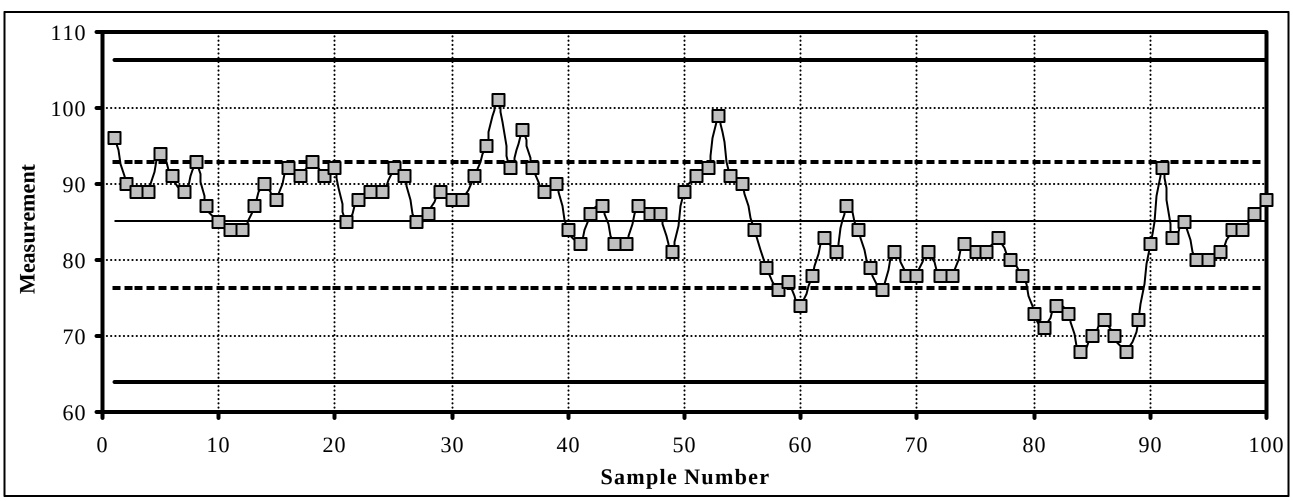

Figure 2: 100 Measurements from a Chemical Process

which are based on estimates of the process mean and standard deviation derived explicitly from the data and the fitted ARMA(1,0) model. (Specifically, the estimated mean and standard deviation are $\mu \approx 85.1$ and $\sigma \approx 7.0$, respectively, for which $\mu - 3\sigma = 64.1$ and $\mu + 3\sigma = 106.1$.)

The wider control limits based on the ARMA(1,0) model are limits within which we would expect the vast majority of future observations to lie in the long run as long as the process remains in control. Their center value, $\mu \approx 85.1$, is the average measurement we would expect to observe in the long run as long as the time series model remains appropriate. In the short run, however, the situation is different. For example, given the history of the measurements up to time 100, the time series model also can be used to develop a forecast of $x_{101} \approx 85.9$ and a 3-sigma prediction interval of $74.7 \leq x_{101} \leq 97.1$ for the measurement expected at time 101. Since this prediction interval is narrower and has a different center value that the "long-run" control limits, we can see that the predictability of an autocorrelated process in the short run can differ substantially from its predictability in the long run.

We can summarize our concerns about the assumption that a process is in control as long as it behaves according to an ARMA(p,q) model as follows. First, this assumption implies that it is all right for an in-control process to wander away from its mean or target. We suspect that many quality managers would have reservations about equating a meandering process with an in-control one. Second, this assumption overlooks the fact that a major reason for maintaining a process in a state of control is to ensure its capability. Although we do acknowledge that an in-control process <u>will</u> eventually wander back to its mean (and thus be capable in the long run), we doubt that many quality managers would be willing to tolerate short-term deviations from the "capable" state as the process meanders. Finally, our reservations multiply when we consider the fact that one generally requires relatively large amounts of data to fit an ARMA(p,q) model to data and agree with Montgomery and Mastrangelo (1999) that doing so may "frequently require more effort than may be justified in practice." On the other hand, it should also be apparent that, if one fails to account for the autocorrelation between measurements, then one's estimates of the process mean, standard deviation, and capability could be grossly misleading, particularly if relatively few observations are used.

A General Method for Monitoring Process Capability

To overcome some of our concerns about the use of time series models to represent the behavior of an in-control, autocorrelated process, we thus suggest the use of a "control" scheme that explicitly tracks process capability. We do so cautiously because this suggests a new management framework for statistical process control wherein control—which traditionally has regarded the in-control state as inherent and internal to the process—is tied explicitly to capability—which is evaluated relative to external requirements.

An implicit basis for our proposed approach is the expansion of the types of causes of variation from two to three:

<u>Assignable causes</u>—those factors that contribute most greatly to observed variation and can be detected, identified, and eliminated.

Common causes—the sources of the inherent variability within a system that are not feasible to detect, identify, or eliminate and manifest themselves as a series of independent and identically distributed random shocks to the system.

Structural causes—the sources of autocorrelation in an in-control process that cannot be eliminated but can be adjusted for by resetting the process to its mean or target value.

Note that structural causes straddle the line between assignable and common causes in that they can be compensated for but not eliminated. Thus, when our system generates an out-of-control signal, a search for an assignable cause should be initiated, and if one is found, it should be eliminated. If not, the process should be reset to its mean or target value.

The capability monitoring system we propose is based on the assumption that behavior of real-world times series (in the short-term, at least) can be reasonably predicted by an ARMA(1,1) model of the form

$x_i = \delta + \varphi_1 x_{i-1} + \varepsilon_i - \theta_1 \varepsilon_{i-1}$,

where (as before) $\varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots$ are as a sequence of independent and identically distributed random variables with mean zero and standard deviation σ and the model parameters δ, φ_1, and θ_1 are appropriate model parameters. (We have found this model to be relatively easy to fit and sufficient for modeling the vast majority of processes cited in the literature, with the exception of those with a strong seasonal pattern.)

Given this assumption, Zalewski (1995) shows that process capability can be effectively monitored by charting a short-term estimate of the process capability index given by

where $\mu_{est,i}$ is a forecast of the next measurement in the time series computed from the i^{th} measurement via

$\mu_{est,i} = \delta + \varphi_1 x_i - \theta_1 e_i$,

$$C_{est,i} = \mathrm{Min}\left\{\frac{USL - \mu_{est,i}}{3\sigma_{est,i}}, \frac{\mu_{est,i} - LSL}{3\sigma_{est,i}}\right\},$$

$\sigma_{est,i}$ is an estimate of the standard deviation of the next random shock to be experienced computed via

$\sigma^2_{est,i} = \lambda e_i^2 + (1 - \lambda)\sigma^2_{est,i-1}$,

and λ is a suitably chosen smoothing constant (λ = 0.015 works well in general), and e_i is the forecast error associated with the i^{th} measurement (an estimate of the random shock experienced at time i – 1), defined by

$e_i = x_i - \mu_{est,i}$.

The preceding formulas require initial estimates $\mu_{est,0}$ and $\sigma^2_{est,0}$, which can be found via

$\mu_{est,0} = \delta/(1 - \varphi_1)$ and $\sigma^2_{est,0} = \sigma^2$.

When the model parameters δ, φ_1, θ_1, and σ^2 are not known and must be estimated from data, well-established methods can be used for their initial estimation [usually by minimizing the sum of the squared residuals from the postulated ARMA(1,1) model; see Box and Jenkins (1976)]. These can then be updated as new data is obtained.

Figure 3 charts the values of our short-term estimate of capability for the measurements plotted in Figure 2, assuming specification limits of USL = 55 and LSL = 115, respectively. (These were chosen arbitrarily so that the specifications would be centered roughly at the process mean and correspond to a long-term process capability index of $C_{pk} \approx 1.4$.) To assess whether or not this process is in capable state, we need to determine appropriate critical values for this statistic. Since we are concerned primarily about poor process performance relative to the specifications, we recommend the use of a single, lower critical limit. In practice, this limit should be chosen to balance the probability of a false alarm against the probability of detecting a shift of a specified magnitude (both of which become larger as the lower limit is decreased.) We have found that values between 0.85 and 1.0 work well, with 1.0 being intuitively attractive since values of the index larger than this place the short-term process mean farther than three standard deviations from the specification limits.

Performance of the Proposed Method

In order to test the proposed capability monitoring system, we ran a series of simulated experiments; two of these are described here to provide a general impression of performance [see Zalewski (1995) for more experiments]. In each experiment, we set (without loss of generality) $\delta = 0$ and $\sigma^2 = 1$. The experiments differed in the values of the parameters φ_1 and θ_1 assumed for the in-control process: the first simulated an uncorrelated process (i.e., one with $\varphi_1 = \theta_1 = 0$) while the second simulated a highly correlated process with $\varphi_1 = 0.95$ and $\theta_1 = 0.45$.

Within each experiment, 1,000 sequences of measurements were simulated to determine the average run length—the average number of measurements required to produce an out-of-control signal—for our method as a function of a specified amount of shift in the process mean or standard deviation. Each simulated sequence was initiated by generating measurements from

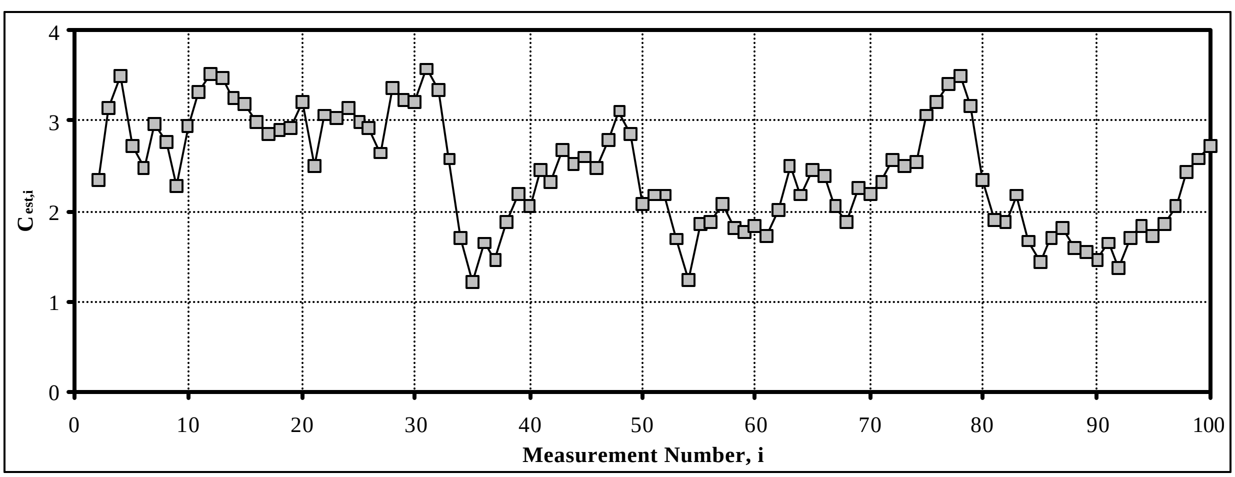

Figure 3: Values of the Proposed Capability Monitoring Statistic

an ARMA(1,1) model driven by simulated standard normal random shocks. In each experiment, the lower control limit was set by trial and error to produce an average run length of approximately 370 (the same as for a standard control chart) when the process is in control.

Tables 1 and 2 display the results of the two experiments. In each table, the proposed method is compared with (i) a time-series-based chart, called a special cause chart (SCC) by Wardell, Moskowitz and Plante (1994), (ii) an X-chart for individual measurements, and (iii) an exponentially-weighted-moving average (EWMA) chart, as also suggested by these authors.

Table 1: Comparison of Average Run Lengths – Uncorrelated Process

Shift in the Process Mean					**Shift in the Process Standard Deviation**				
Mean shift in multiples of σ	Method				Increase in σ (percent)	Method			
	C_{est}	SCC*	X*	EWMA*		C_{est}	SCC	X	EWMA
0.0	380.	370.	370.	369.	0	380.	370.	370.	371.
0.5	48.8	155.	155.	**28.2**	25	**41.0**	61.0	61.0	99.8
1.0	14.5	43.9	43.9	**9.7**	50	**19.4**	22.0	22.0	46.8
1.5	8.4	14.9	14.9	**5.8**	100	9.7	**7.5**	**7.5**	19.3
2.0	5.8	6.3	6.3	**4.2**	200	4.9	**3.2**	**3.2**	8.0
2.5	4.4	3.2	3.2	**3.3**	300	3.2	**2.2**	**2.2**	4.9
3.0	3.3	2.0	2.0	**2.8**					

*ARL as reported by Wardell, Moskowitz and Plante (1994).

Table 2: Comparison of Average Run Lengths – Correlated Process

Shift in the Process Mean					**Shift in the Process Standard Deviation**				
Mean shift in multiples of σ	Method				Increase in σ (percent)	Method			
	C_{est}	SCC	X	EWMA		C_{est}	SCC	X	EWMA
0.0	380.	370.	370.	369.	0	380.	370.	370.	n/a
0.5	253.	250.	263.	**232.**	25	74.	**61.**	107.	n/a
1.0	123.	275.	109.	**94.**	50	28.	**22.**	49.	n/a
1.5	**46.**	148.	51.	**46.**	100	10.	**7.**	20.	n/a
2.0	**14.**	43.	21.	24.	200	4.5	**3.2**	7.2	n/a
2.5	**3.9**	6.6	7.3	13.5	300	3.0	**2.2**	4.2	n/a
3.0	1.5	**1.3**	2.2	8.8					

Somewhat surprisingly, the proposed method is competitive with an EWMA chart for detecting most types of shifts in an uncorrelated process (for which the EWMA is designed). Similarly, for the correlated process, the proposed method is competitive to the other methods for every type of shift tested and is superior in at least three cases. Overall, this suggests that the proposed approach is at least a reasonable alternative to the other control procedures.

Conclusion

As stated earlier, our goal was to initiate discussion among managers of technology with regard to the merits of a nontraditional approach for monitoring process capability in processes that produce autocorrelated measurements. We attempted to establish this approach by appealing to fundamental concepts of statistical process control and presented arguments for its reasonableness. We now invite feedback from the quality management community.

References

Alwan, L. C. and Roberts, H. V. (1989). "Time Series Modeling for Statistical Process Control." *Statistical Process Control in Automated Manufacturing.* J. B. Keats and N. F. Hubele, eds.,

Marcel Dekker, New York, NY.

Atienza, O. O., Tang, L. C., and Ang, B. W. (1998). "An SPC Procedure for Detecting Shift Levels of Autocorrelated Processes," *Journal of Quality Technology*, 30 (4), pp. 340-351.

American Society for Quality Statistics Division (1996). *Glossary and Tables for Statistical Quality Control*, 3rd ed., Quality Press.

Berthoux, P. M., Hunter, W. G., and Pallsen (1978). "Monitoring Sewage Treatment Plants: Some Quality Control Aspects." *Journal of Quality Technology*, 10 (3), pp. 179-193.

Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control,* rev. ed., Holden-Day, Inc., San Francisco, CA.

Faltin, F. W. and Woodall, W. H. (1991). "Discussion." *Journal of Quality Technology*, 23 (3), pp. 194-197.

Montgomery, D. C. (1996). *Introduction to Statistical Quality Control,* 3rd ed., John Wiley & Sons, New York, NY.

Montgomery, D. C. and Mastrangelo, C. M. (1991). "Some Statistical Process Control Methods for Autocorrelated Data." *Journal of Quality Technology*, 23 (3), pp. 179-193.

Wardell, D. G., Moskowitz, H., and Plante, R. D. (1994). "Run-Length Distributions of Special-Cause Control Charts for Correlated Processes." *Technometrics*, 36, pp 3-17.

Yourstone, S. A., and Montgomery, D. C. (1989). "Development of a Real-Time Statistical Process Control Algorithm." *Quality & Reliability Eng. International*, 5, pp. 309-317.

Zalewski, D. J. (1995). *Methods for Monitoring Process Control and Capability in the Presence of Autocorrelation*. Ph.D. dissertation, Air Force Institute of Technology.

Title: A Management Framework for Process Control and Capability
In The Presence Of Autocorrelation

Daniel J. Zalewski, Ph.D., Office of the Secretary of Defense

Dan Zalewski is a graduate of the United States Air Force Academy who received his Ph.D. from the Air Force Institute of Technology in 1995. Since then he has been an analyst for the Air Force at the Pentagon and is currently assigned to the Office of the Secretary of Defense, Program Analysis and Evaluation, serving as the Director of the Simulation and Analysis Center.

Edward F. Mykytka, Ph.D., University of Dayton

Ed Mykytka is an Associate Professor in the Department of Engineering Management and Systems at the University of Dayton, his undergraduate alma mater. He received a Ph.D. in Systems Engineering from the University of Arizona in 1983. His research interests include discrete-event system simulation (particularly random variate modeling and generation), quality improvement and statistical process control, stochastic modeling, and applied statistics.