Forecasting and Dynamic Control of SAP and Engineering Process Adjustment Integration

S. Y. Gajjal[1], Dr. A. P. S. Gaur[2], Dr. S. R. Kajale[3]

[1]Asst. Professor, SCOE, Vadgaon (BK)

[2]Professor & Head, B.I.E.T,

[3]Director, S.G.G.S Engg. & Tech,

Abstract – Industrial Statisticians frequently face problems in their practice where adjustment of a manufacturing process is necessary. In this paper, a view of the origins and recent work in the area of Statistical Process Adjustment (SPA) is provided. A discussion of some topics open for further research is also given including new problems in semiconductor manufacturing process control. The goal of the paper is to help display the SPA as a research area with its own identity and content and promote further interest in its development and application in industry.

Key Words – Engineering Process Control, Time Series Control, EWMA controllers,

--◆--

1. INTRODUCTION

Located at the intersection of Control Theory, Time Series Analysis, and Statistical Process Control, the Statistical Process Adjustment (SPA) field is a set of Statistical Techniques aimed at modeling, and hence, forecasting and controlling a dynamic process. Two distinctive characteristics of SPA are a) that the process responses relate to quality characteristics of a product (or of a process producing it), and b) the implementation of the adjustments is not fully automatic since SPA corresponds to a higher-level supervisory controller, i.e., a controller of lower-level controllers which in turn operate on a production process. Property a) differs from many control theory applications where some physical variable is of interest, but the aim is not necessarily *quality* control, and b) emphasizes the hierarchical nature with which the adjustments are implemented, on whole complex processes or machines made of several different components, but modeled as a single processing stage. Given the complexity of the machine or process, only a statistical, i.e., data-based modeling is feasible. This is in contrast to first principles models frequently used in control theory.

A key question we would like to address in this paper is: is SPA an area with enough intellectual content and practical relevance to justify its study within Statistical methodology?

We pose this question because two widespread conceptions found among Statisticians and Engineers:

1. process adjustments are, for the most part, unnecessary in practice. This believe is based mainly on statements in Deming's writings and in particular in relation to his \funnel experiment";

2. process adjustments are of course necessary, but practically all the relevant problems have been solved by control theorists. While control theory is a fertile and active area of research, all the problems in SPA have by now been solved. Most of the work on SPA is simply a repetition of previous control theory work.

Believe 1 is found mainly among Statisticians and has been discussed in the literature at length based on the funnel experiment (e.g., MacGregor 1990); believe 2 is found among Engineers, and has not been discussed much, if at all, in the literature. It is the purpose of this paper to show that both viewpoints are misconceptions, but we will place more emphasis on arguing against the second viewpoint. In order to do this, we will review how the SPA field originated, what were the main initial problems, how it has evolved in general terms, and perhaps more importantly, what recent work relevant in industrial practice has been conducted. No effort to provide

a complete literature review was made. For bibliographic references up to 2001 see Del Castillo (2002a).

It is hoped that the problems described here will provide renewed impetus to the area. The paper closes with a discussion of relevant, practically-important problems which are still open for solution. The objective is to provide, by example, enough evidence for an unqualified \yes" answer to the key question posed above, and provide some hints for further research.

2. ORIGINS OF SPA

Since its origins in the early 60's, work in what we now can consider the SPA field was developed by both Control Engineers (working in quality control applications) and by Statisticians, who, for the most part, had a background in Chemical Engineering. This is more than a simple anecdote, since it determined what type of processes and corresponding problems were initially studied in this field. SPA originated from work apparently done independently by Box and Jenkins (1962, 1963) on adaptive optimization and control and minimum mean square error (MMSE) control, and by °AstrÄom (1963) on \minimum variance" (equivalent to MMSE) control (°AstrÄom was interested in implementing Kalman's ideas on Adaptive Control based on operating data). While there were some interesting papers on process control (as opposed to control charting) written by Statisticians in the late 50's and early 60's (e.g. Barnard, 1959), the work by Box and Jenkins was the most influential in the Statistics literature.

The MMSE control problem relates to finding a rule (a \controller") that tells us how to vary a controllable factor x_t such that the MSE of a dynamic response y_t (which we assume to be deviations from target) is minimized in the following transfer function model:

$$y_t = \frac{B'(\mathcal{B})}{A'(\mathcal{B})}x_{t-k} + \frac{C'(\mathcal{B})}{D'(\mathcal{B})}\varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} (0,\sigma^2). \tag{1}$$

Here, B^1; A^1; C^1 and D^1 are polynomials in the backshift operator B (all polynomials start with a one except B^0 which starts with some arbitrary constant b_0) and k is the delay, i.e., the number of whole discrete time periods between the controllable factor is changed and its effect on the response appears to be observed for the ¯rst time. It is assumed all-time series have equidistant observations in time. Even control engineers to this day call this model the *Box-Jenkins model*. Contrary to other different ways of writing a transfer function model,

(1) has a natural \signal plus noise" interpretation, since if $x_t = 0$ for all t, then is an ARIMA model that represents the uncontrolled output (here polynomial D^1may have one or

more roots on the unit circle, allowing to model *homogeneous non-stationarity*, see Box, Jenkins and Reinsel, 1994). This also occurs if x_t =constant since then the ¯rst term on the right is simply a constant. Model (1) also has the advantage of avoiding multiple common terms when fitting (Box et al., 1994).

Despite the nice interpretation and model-fitting advantages, it is perhaps easier to de-rive an MMSE controller with fewer polynomials around. The ARMAX form of a transfer function model, used by AstrÄom, is

$$A(\mathcal{B})y_t = B(\mathcal{B})x_{t-k} + C(\mathcal{B})\varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} (0,\sigma^2) \tag{2}$$

with the obvious relations between the two models being $A = A^0D^0$, $B = B^0D^0$, and $C = C^0A^0$. Let n be the maximum order of the polynomials A, B and C. The optimal MMSE.feedback controller, found by both AstrÄom and Box and Jenkins is

$$x_t = -\frac{G(\mathcal{B})}{B(\mathcal{B})F(\mathcal{B})}y_t \tag{3}$$

From this initial work, an explosion of related work took o®. °AstrÄom and his colleagues and students (notably L. Ljung) went ahead and founded the Swedish school of Adaptive Control. Adaptive Control is by now a mature discipline within Control Theory (°AstrÄom and Wittenmark, 1989). Adaptive controllers continuously re-estimate the parameters of a given model, thus their properties are di±cult to analyze. Although such recursive estimators are known to *burst* if the inputs (the x's) do not vary enough, in practice several safeguards that monitor the \health" of the estimator (not unlike SPC schemes) are applied to provide persistent excitation without bursting (see Ljung, 1999) (the regression equivalent of the lack of excitation problem is an X^0X matrix very ill-conditioned due to similar rows in X).

For their part, Box and Jenkins and their students (notably J. MacGregor) continue to develop SPA in the 70's and beyond by studying problems with a clear Statistical content. We will review some of this work. But ¯rst it is pertinent to address a not so-well-known misconception created by Deming's funnel experiment.

3. WHEN ARE PROCESS ADJUSTMENTS NECESSARY?

There is a considerable good understanding of when adjustments are necessary and why, see e.g., MacGregor (1988) and Del Castillo (2002a, chapter 1), who discuss this issue based on Deming's funnel experiment. We would like to point out a common misconception made by some authors who prefer not to contradict directly Deming's

remarks, as expressed, e.g., in his *Out of the Crisis* book (Deming, 1982). It is sometimes argued that Deming's remarks about not to adjust are actually correct provided the process mean is not moving. It is argued that if the process mean changes with time, then adjustments are needed, otherwise, they are not. This is incorrect. *A moving mean is neither a necessary nor a su±cient condition for adjustments to be required.* To show why, consider the funnel experiment in which, the analogous univariate process would obey Shewhart's model (the marbles evidently obey a bivariate process):

$$y_t = \mu + \varepsilon_t, \quad \varepsilon_t \stackrel{iid}{\sim} N(0, \sigma^2), \quad t = 1, 2, \dots \quad (4)$$

where y_t is the observed deviation from target and μ is the mean deviation from target. Deming assumes the funnel to be on target, i.e., $\mu = 0$. There are two important aspects to notice: first, the process starts on target and remains there unless an adjustment is made. Second, the observations form an i.i.d. (normal if the ε_t's are normal) sequence.

In a very important but somewhat neglected paper (despite being reprinted in 1983), F. Grubbs (1954) assumed the second condition above but assumed that, at startup of the process, $\mu = d \neq 0$, where d is a setup error. If the only cost of interest is the cost incurred when running the process offtarget, then it is evident that corrective action is necessary. How to do this in such a way that is minimized is called the *Setup Adjustment Problem*, with many variants that have been studied considerably in the last 7 years (Del Castillo, 1998, Trietsch, 1998, 2000, Pan and Del Castillo, 2003, 2004, Colosimo et al., 2004, Lian et al., 2005).

Note how the process mean is *constant*, there is no autocorrelation, and still process adjustments are needed. This shows by counterexample that it is not *necessary* that the process mean moves for adjustments to be required. A moving mean is neither a *sufficient* condition for adjustments to be required: as a counterexample consider the case of a process with a moderate drift in the mean such that all product will be within specifications (or not too far from target to cause a substantial cost) for the duration of the production run, and suppose the cost of adjustments is relatively very large. Then, it follows that adjustments are not justified.

In conclusion, the need for process adjustments depends on the process model and the cost structure. Evidently, if all conditions behind Deming's funnel experiment hold, then simple process monitoring is optimal from a MMSE point of view.

4.1 The Setup Adjustment Problem

Solutions to the setup adjustment problem and to many of its variants, some to be described shortly, are very important for the control of *discrete part* manufacturing processes. In this type of processes, the operation of setting up a machine for production of a new lot may induce offsets or shifts in the values of the quality characteristics of the parts relative to their targets. No disturbance other than the setup offset and white noise are assumed. If the unknown offset is a constant d, the deviation from target at time t can be expressed as (Del Castillo et al., 2003):

$$y_t = \mu_t + v_t, \quad v_t \stackrel{iid}{\sim} (0, \sigma_v^2) \quad (5)$$

$$\mu_t = \mu_{t-1} + \nabla x_{t-1}, \quad t = 2, 3, \dots \quad (6)$$

$$\mu_1 = d + x_0 \quad (7)$$

Del Castillo et al. (2003) show how (9) results from using a simple Kalman filter to esti-mate the \state" μ_t and adjusting by b_t. This formulation allows to apply Linear Quadratic Gaussian theory to extend the basic setup adjustment problem to multiple input-multiple output (MIMO) problems, problems with errors in the adjustments, and problems with quadratic adjustment costs. It also allows to make interesting connections between the setup adjustment problem and other Statistical techniques such as Stochastic Approximation and Recursive Least Squares. These extensions and connections were not possible using Grubbs' more complex approach to the problem.

While the basic setup adjustment problem can be solved with well-established control theory techniques, many more important variations are problems that cannot be solved making use of existing control techniques and require new methodology.

$$\nabla x_t = -\frac{1}{t} y_t \quad (8)$$

$$\nabla x_t = -\frac{1}{t + \frac{\sigma_v^2}{\sigma_d^2}} y_t. \quad (9)$$

From a Statistical perspective, the most interesting and practical variation of this problem is when the process parameters μ_d, σ_d^2, and σ_v^2 are unknown. Then the problem is an Adaptive Control problem, as it involves controlling a process with unknown parameters. The structure of the problem, however, has not been addressed by Control theorists, as far as we know, since here the variances need to be estimated on -line.

Recent work in setup adjustment under unknown paraments is Bayesian and based on Markov Chain Monte

Carlo (MCMC) and Sequential Monte Carlo (SMC) techniques. In Colosimo et al. (2004), the assumed model is analogous to (5-7):

$$y_{ij} = \theta_{ij} + v_{ij} \tag{10}$$

$$\theta_{ij} = \theta_{i(j-1)} + \nabla x_{i(j-1)} \tag{11}$$

$$\theta_{i0} \sim (\mu, \sigma_\theta^2), \quad v_{ij} \overset{iid}{\sim} (0, \sigma_v^2) \tag{12}$$

The Bayesian MCMC controller learns how to \anticipate" the offsets, providing a perfor-mance that eventually mimics a feed forward controller. Unnecessary adjustments that may inflate the overall process variance are reduced via a conditional ¯rst adjustment rule, in which (13) is implemented only when a credibility interval for ¹ excludes zero. See Lian et al. (2005b) for details.

Recent work on setup adjustment includes the case when parameters are known with sufficient accuracy. If this is the case, then the sum of the total cost of running the process o® target and of adjusting can be minimized by defining a *schedule* of adjustments, much in the sense of a maintenance plan. See Trietsch (2000) and Pan and Del Castillo (2004). Other recent work includes the case of an asymmetric o®-target cost, a common situation in discrete part manufacturing. One approach is to let the process converge to target from the side of least cost. Stochastic approximation techniques can then be used for this purpose. See Colosimo et al. (2005) for more details.

Other relevant variations of the setup adjustment problem, in particular, integration with process monitoring, the case when there a re fixed adjustment costs {resulting in \dead band" policies{ and the use of SMC techniques are described in the next sections.

4.2 Dead band" adjustment policies

The dynamic programming problem, solved by Crowder, yields a dead band solution analogous to the infinite-horizon solution but with dead band limits L_t that \funnel out" as the end of the production run approaches. The implication is that if the process will end soon, an adjustment at that point brings less future benefits than an adjustment early in the production of the lot. Jensen and Vardeman (1993) consider the same finite-horizon problem as in Crowder (1992), but studied the case when adjustment errors can occur randomly. They show that even if no fixed adjustment cost exists, adjustment errors imply a dead band policy.

The work on dead band adjustment thus far summarized is based on knowing all process parameters. For the model

assumed by Crowder (equivalent to that used by Box and Jenkins, equation 14):

$$y_t = g x_{t-1} + \frac{1 - \theta\mathcal{B}}{1 - \mathcal{B}} \varepsilon_t. \tag{13}$$

$$\nabla x_t = -\frac{1}{g} \widehat{y}_{t+1|t}, \quad \widehat{y}_{t+1|t} = (1-\theta)y_t + \theta\widehat{y}_{t|t-1} \tag{14}$$

Recently, Crowder and Eshleman (2001) propose an alternative to the Bayesian SMC approach just described based on using Maximum Likelihood Estimation for the variances from a set of *n open loop* runs (i.e., data obtained when the controller is disconnected), and then plug-in those estimates in the usual Kalman filter estimate of the state. They reported small sample properties of the estimators, concluding that at least between 25 to 50 observations are necessary to obtain reliable estimates. It would be interesting to see how can the MLE method be modified for closed-loop data, and how it behaves compared to the Bayesian approach. It seems likely that for non-informative priors, the performance should be quite close, with the advantage of the Bayesian method of being able to incorporate prior information about the parameters in case it exists, which would improve convergence of the estimation process.

It is important at this point to emphasize that these and other variations of dead band control are not considered in the control theory literature.

4.3 PI and EWMA control

The last term, with action proportional to the difference of the response, is frequently not used in practice (Del Castillo, 2002a, Chapter 6). This results in PI controllers, which have received considerable attention in the Statistics community due to the work by Box and Luceno (1997). These authors show how to implement a PI controller graphically (an idea ¯first shown by Box and Jenkins, 1963), and show how PI controllers are quite robust with respect to variations in the assumed process model. They convincingly show that the inflation in variance due to adjusting with a PI controller a process that requires no adjustment, {like Deming's funnel{ is quite moderate.

$$x_t = K_P y_t + K_I \sum_{i=1}^{t} y_i + K_D \nabla Y_t \tag{15}$$

$$\nabla x_t = K_P \nabla Y_t + K_I Y_t + K_D \nabla^2 Y_t. \tag{16}$$

The robustness of PI controllers is widely acknowledged and known by process engineers in practice. Process engineers working in industry know well the adagio that

says that all it takes most of the times is a good integral controller.

The robustness of PI controllers comes from its integral action, controller, one in which the controller linearly tracks the response, is a quite poor controller in general, since, for example, it does not provide offset-free control. In contrast, and this should be of interest to persons familiar with Statistical Process Control (SPC) charts, an integral controller will compensate against shifts in the mean of a stationary process. This will make detection of a shift in a PI-controlled process difficult (see Jiang and Tsui, 2002, and the discussion below on SPC-EPC integration). The time to recover of the process will be a function of the shift magnitude and of the integral parameter K_I . In principle, if the adjustments are unconstrained in size, and integral controller will eventually compensate against shifts of *any* size. The design of a PI controller consists in selecting K_P and K_I (see Box and Luceno, 1997, Tsung et al., 1998). Constrained input variance PI controllers are discussed by Box and Luce~no, 1997. An input variance constrained PI controller that tunes K_P and K_I on-line (i.e., it is self-tuning) was developed by Del Castillo (2000). The approach solves for the Lagrange multiplier of the constraint Var(rx_t) = c, and uses this multiplier in the Clarke et al. (1975) controller, which utilizes it to constrain the input variance (see Del Castillo, 2002a). Since it is not based on a recursive estimator and utilizes Box and Luceno suggested settings for the parameters as initial values, the bursting behavior typical of adaptive controllers is avoided.

A minimal condition for a good controller is that it must be stable. Stability has not always been considered in the SPA literature, as some of the discussants of the Box and Jenkins (1962) paper pointed out. Stability conditions of EWMA and DEWMA controllers has been a matter of study in the last 10 years (see, e.g., Ingolfsson and Sachs, 1993, Guo et al., 2000, Del Castillo, 1999). For a large variety of disturbances, an EWMA controller is stable if and only if $j1$ i ¸ $»j < 1$, where $» = g=gb$. Stability conditions for DEWMA controllers with unit delay were derived by Del Castillo (1999) and later simplified by Tseng et al. (2002). They show how if N_t is an ARIMA(p,d,q) with drift model ($d \cdot 2$), a sufficient condition for asymptotic stability is that $g=gb < 3=4$. MIMO double EWMA controllers have been studied by Del Castillo and Rajagopal (2002).

The work thus far described on PI and EWMA control is just an application of existing methodology in Control Theory. To the eyes of control theorists, this work looks as straight-forward approaches, compared to the complexities of current control theory research. Let us now turn to some new methodological developments that built on the previously cited work. In the last section of this paper we will also delineate some further problems related to EWMA

control that arise in practice and require new methodologies.

Some interesting recent work by Hamby et al. (1998) introduces the concept of prob-ability of stability" and \probability of performance" in the design and analysis of EWMA controllers. These authors noted how in run to run applications, the gain g is usually fitted off-line based on designed experiments. In their paper, the gain is actually a vector $µ$, as they analyzed the multiple input, single output case (MISO) model:

The aim is the same, to develop controllers that are insensitive to uncertainties in the assumed model. As mentioned by Apley and Kim (2004), Robust Control, and in particular, H_1 optimization is a mature field that has dominated most of Control Theory research in the last couple of decades. Its central precept is that if one can place deterministic bounds on the unknown parameters of a process, then a worst case performance index which considers variations of the parameters within such bounds can be optimized and a robust controller design obtained.

In the type of manufacturing quality control applications where SPA has evolved, such view of robustness is not satisfactory since parameters are usually estimated from production data of complex industrial processes (this is echoed by °AstrÄom and Witten mark, 1989, when proposing Adaptive Control techniques). In such environment, it will be hard or impossible to place definite bounds on the variation of the parameters. These noisy, data rich environments imply that a probabilistic measure of uncertainty will generally be possible and preferable. The means by which probabilistic measures can be developed, as in the last two paragraphs above, is Bayesian inference, to which we return in Section 5.

4.4 SPC-EPC" integration

"Fault Detection" and "Advanced Process Control" in semiconductor manufacturing circles, the integration of SPC tools and "Engineering Process Control" methods to operate on the same process is a problem that naturally falls within the SPA field. There have been two fundamentally different approaches for doing this integration of techniques:

1.   The SPC mechanism acts in conjunction with an MMSE, PI, or other known controller which is active all the time. This approach was stated conceptually by Vander Weil et al. (1992), Faltin et al. (1993) and Tucker et al. (1993) who coined the term Algorithmic Statistical Process Control". In this case, monitoring is typically conducted on the output of a controlled process, although approaches have been proposed to monitor both x_t and y_t jointly (Tsung

and Shi, 1999). Note how the cause of a SPC signal can be assignable to a faulty feedback controller. In this way, the SPC scheme helps to monitor both the health of the process and of the EPC scheme. This has connections with the considerable body of work on SPC for auto correlated data (literature too numerous to cite here), since the output of a controlled process is typically correlated in time (e.g., consider the closed-loop equation 4). This, in turn, relates to the analysis of the response patterns or \signatures" of a dynamic system to specific upsets (see, e.g., Yang and Makis (2000), Tsung and Tsui, 2003).

2. The SPC mechanism acts as a trigger of the EPC mechanism. This is the approach of authors such as Sachs and Ingolfsson (1995) and of Guo et al. (2000) in the area of \run to run" control (an early reference of this approach is Bishop, 1965). Usually, a step-like disturbance is assumed to occur with some probability. An SPC-like scheme is from a MSE point of view as p increases. Box and Luce~no (1997, Chapter 5), showed that the IMA(1,1) model is in fact a good approximation to the random jump model even for relative low values of p. Despite this fact, if one always uses an EWMA controller based on an IMA(1,1) model, monitoring for the eventual elimination of assignable causes will be a task harder to accomplish.

An instance of recent work along the first line of reasoning described above is by Jiang and Tsui (2002), who studied the Average Run Length properties of SPC charts designed to monitor the type of auto correlated processes which result from adjusting a process with a MMSE or PI controller. The case of an MMSE controller is particularly tractable, given the closed-loop equation (4), thus essentially the problem is one of monitoring a MA($k_{\,\textrm{\scriptsize i}}$ 1) process. They concluded that for PI-controlled process, detecting the presence of a shift is difficult by monitoring the output y_t. They suggested instead to monitor the level of the controllable factor (x_t). This can actually be generalized to any controller that has integral action: the integral action will compensate for the shift disturbance, thus only a transient \spike" in y_t will appear. The more aggressive the integral action is, i.e., the larger K_I is, the shorter this window of opportunity to detect will be. In some industrial processes, e.g., semiconductor manufacturing, aggressive I control is common, so this is a relevant problem in practice. Because of this \masking" of the assignable causes that impede their removal through the usual {but not modeled { process improvement steps that SPC recommends (called \technical feedback" by Box and Jenkins, 1962), some authors have argued against process adjustments. This is typically not an option, particularly if the process drifts, i.e., if it is open-loop unstable.

Thus, in this alternate SPC/EPC integration approach, an EPC mechanism is invoked only when it is necessary. This

alternative resembles a dead band controller and the \machine tool" problem, but the assumed disturbances and motivations are different. The machine tool problem assumes an IMA(1,1) disturbance, which is well-known to be optimally forecasted through the EWMA in eq. (15). There, the fixed cost of adjusting implies the dead band structure of the solution; in the integrated SPC-EPC approaches the SPC acts as a dead band since no other disturbance is supposed to exist between shift detection times. Interestingly, as p, the probability of a shift in any time point, increases, then the corresponding stochastic process for y_t increasingly resembles an IMA(1,1) process (which in itself can be thought of as a random walk observed with error). This implies that when p is large, simply using an EWMA controller based on (15) without a dead band will work better from a mean squared deviation point of view than the type of integrated CUSUM/harmonic rule described here. This was also noted by Chen and Elsayed (2002), who studied how to tune an EWMA controller $x_t = {}_{\textrm{i}}a_t{=}g$ where a_t is an EWMA of the y_t's, when the disturbances follow the step-like process described by (17). From our description above, the IMA(1,1) will be an increasingly better model and a controller based on it will be increasingly closer to optimal

In the type of manufacturing quality control applications where SPA has evolved, such view of robustness is not satisfactory since parameters are usually estimated from production data of complex industrial processes (this is echoed by °AstrÄom and Witten mark, 1989, when proposing Adaptive Control techniques). In such environment, it will be hard or impossible to place definite bounds on the variation of the parameters. These noisy, data rich environments imply that a probabilistic measure of uncertainty will generally be possible and preferable. The means by which probabilistic measures can be developed, as in the last two paragraphs above, is Bayesian inference, to which we return in Section 5.
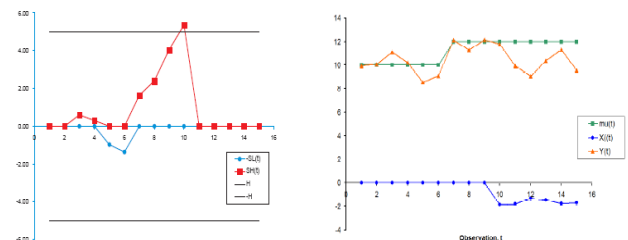

Figure 1: CUSUM-harmonic rule integrated approach. Left: CUSUM chart; right: observed and mean quality characteristic (top) and controllable factor values (bottom).

Note how the \machine tool" dead band does not include a monitoring scheme, thus assignable casue *removal* is not

possible using a dead band scheme (despite its resemblance of an SPC chart).

5. BAYESIAN METHODS IN PROCESS ADJUSTMENT

We have already referred to recent SPA methods that are Bayesian, such as setup adjustment using MCMC techniques (Colosimo et al., 2004), dead band schemes for process adjustments (Lian and Del Castillo, 2005), and cautious control (Apley and Kim, 2004). In this section we further comment on the potential of modern Bayesian statistical techniques in SPA and some areas open for research. Well-known control theory techniques have connections with Bayesian techniques or can be interpreted in a Bayesian way, two examples being Kalman filtering for state estimation (with known parameters) and Adaptive Control. The main potential for new Bayesian SPA methods, much along the type of problems discussed earlier, is on the adjustment of short-run processes with unknown parameters.

Breakthroughs in numerical integration developed over the last 15 years can now be routinely utilized for posterior inference when non-conjugate priors are desired. In particular, MCMC methods (Gelman et al., 2003) have been developed intensively and proved to provide solutions to previously un tractable problems.

For a problem in which data arrives sequentially in time, however, MCMC methods may not be the best choice. In MCMC, Markov Chains iterations yielding the target posterior distribution are repeated from scratch every time a single new observation y_{t+1} is obtained, without reusing the posterior distribution previously obtained a period before, i.e., at period t. An alternative to MCMC is Sequential Monte Carlo (SMC) methods (see Figure 2). SMC methods also rely on Monte Carlo algorithms for the solution of Bayesian inference problems. In SMC, posterior distributions of \particles" $\mu^{(i)}$ (values of the parameter) are created numerically from calculating associated weights w_i. These weights are recomputed after each observation is obtained based on the likelihood of the corresponding particle given the new datum and the previous set of weights, keeping in this way information from the previous step. The weights are then used to provide posterior estimates of any function of the parameter of interest at time $t + 1$. A major advantage of SMC techniques is that they are considerably faster than MCMC, allowing for on-line control. A brief sketch of the computations required to approximate the expectation of some function of an unknown parameter μ at step i is as follows:
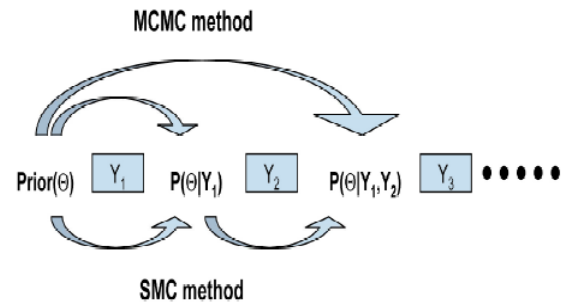

Figure 2: The update of a posterior distribution in MCMC and SMC.

Here $L(\mu^{(j)}jy_i)$ is the likelihood function of the jth particle given the latest observation. If, e.g., interest is in the sample mean, then f ´ 1. The *rejuvenation step* is executed if the sample of particles is too poor. This will tend to happen when $¼(\mu)$ is a non-informative prior. In such case, many particles will be unlikely given the data, so their weights w_i will be zero after a few iterations. The distribution of the w_i's will contain only a few non-zero weights, and will provide biased estimates. A rejuvenation step (Balakrishnan and Madigan, 2004) smoothes the posterior distribution of the particles. Then, importance re sampling of the parameters μ is performed using the updated weights. See Doucet et al. (2001) for more details.

Lian et al. (2005) apply the SMC method to the setup adjustment problem for un-known parameters. They show how SMC gives results equivalent to MCMC but a fraction of the computing time. The bayesian dead band adjustment scheme mentioned earlier (Lian and Del Castillo, 2005) also utilizes SMC. There is a wide range of other relevant control problems with unknown parameters that could be approached with SMC techniques. This includes adaptive filtering problems and, in general, State-Space models. The SMC procedure provides posterior parameter distributions of any relevant parameters which in turn can be used to minimize a variety of cost functions. The solutions so obtained will in general be suboptimal since the certainty equivalence principle (which indicates when using plug-in estimates leads to optimal solutions, see Del Castillo, 2002a, Appendix 8B) applies only in restrictive cases. Nevertheless, the solutions obtained may still have excellent performance considering that a \dual control" optimal solution is computational prohibitive in most cases (see ˚AstrÄom and Wittenmark, 1989). In a given application of SMC to process adjustment, additional work is needed to quantify its performance over some known lower bound or reference point of performance.

A second area we would like to highlight where Bayesian inference can play an important role is in Closed-loop identification. If lack of identifiability is a problem due to not

having enough information about the process parameters, it seem natural to use a Bayesian approach in which any prior information available can be incorporated. How to determine such prior(s) and what type of additional pieces of information one should be able to model with the priors are questions open for future research. MCMC methods used for open and closed loop identification have recently been mentioned by Ninness et al. (2002) and Thil and Gilson (2004), respectively.

CONCLUSION

Process adjustments are, for the most part, unnecessary in practice. This believe is based mainly on statements in Deming's writings and in particular in relation to his funnel experiment", process adjustments are of course necessary, but practically all the relevant problems have been solved by control theorists. While control theory is a fertile and active area of research, all the problems in SPA have by now been solved. Most of the work on SPA is simply a repetition of previous control theory work. In this paper, a view of the origins, present status, and a discussion of some areas for further research on Statistical Process Adjustment methods was given. The goal was to provide convincing examples that would demonstrate the intellectual and practical value of this field of Industrial Statistics, and to promote interest for further research.

REFERENCES

1. Luce~no, A. (2003). \Dead-band Adjustment Schemes for On-line Feedback Quality Control", in Handbook of Statistics, Vol. 22, R. Khattree and C.R. Rao, eds., Elsevier Science B.V.

2. Luce~no, A., and Gonz¶alez, F.J., (1999). \E®ects of Dynamics on the Properties of Feedback Adjustment Schemes with Dead Band". *Technometrics*, 41, pp. 142-152.

3. MacGregor, J. F., (1988). \On-Line Statistical Process Control", *Chemical Engineering Progress*, October, pp. 21-31.

4. MacGregor, J.F., 1990. \A Di®erent View of the Funnel Experiment," Journal of Quality Technology, 22, pp. 255-259.

5. MacGregor, J.F., and Fogal, D.T. (1995). \Closed-loop identi¯cation: the role of the noise model and pre¯lters", *Journal of Process Control*, 5(3), pp. 163-171.

6. Milliken, G.A. and Johnson, D.E. (1984). *Analysis of Messy Data*. New York: Van Nostrand Reinhold.

7. Morari, M., and Za¯riou, A. (1989). *Robust Control*. Prentice Hall, Englewood Cli®s, NY.

8. Moyne, J., Del Castillo, E., and Hurwitz, A., eds., (2000). *Run to Run Process Control for Semiconductor Manufacturing*, CRC Press.

9. Ninness B., Henriksen, S., and Brinsmead, T. (2002). \System Identi¯cation via a Com-putational Bayesian Approach", *Proceedings of the 4st IEEE Conference on Decision and Control*, Las Vegas, NE, pp. 1820-1825.

10. Pan, R., and Del Castillo, E. (2001). \Identi¯cation and Fine Tuning of Closed-loop Pro-cesses under Discrete EWMA and PI Adjustments", *Quality and Reliability Engineering International*, 17, pp. 419-427.

11. Pan, R., and Del Castillo, E. (2003). \Integration of Sequential Process Adjustment and process Monitoring techniques," *Quality & Reliability Engineering International*, 19,4 pp. 371-386.

12. Pan, R. and Del Castillo, E. (2004). \Scheduling Methods for the Setup Adjustment prob-lem," *International Journal of Productions Research*, 41,7, pp. 1467-1481, (2003). Correc-tion, 42(1), pp. 211-212.

13. Sachs, E., Hu, A., and Ingolfsson, A., (1995). \Run by Run Process Control: Combining

14. SÄoderstrÄom, T., Ljung, L. and Gustavsson, I. (1976). \Identi¯ability conditions for linear multivariate systems operating under feedback", *IEEE Transactions on Automatic Control*, 21(6), pp. 837-840.

15. Thil, S. and Gilson, M. (2004). \Closed loop identi¯cation: a bayesian approach", Working paper, Centre de Recherche en Automatique de Nancy (CRAN).

16. Trietsch, D., (1998). \The Harmonic Rule for Process Setup Adjustment with Quadratic Loss", *Journal of Quality Technology*, 30, 1, pp. 75{84.

17. Trietsch, D. (2000). \Process setup adjustment with quadratic loss". *IIE Transactions*, 32(4), pp. 299-307.

18. Tseng, S.-T., Chou, R.-J., and Lee, S.-P. (2002). \Statistical Design of double EWMA Con-troller". *Applied Stochastic Models in Business and Industry*, 18, pp. 313-322.

19. Tseng, S.-T., and Hsu, N.-J. (2005). \Sample Size Determination for Achieving Asymptotic Stability of a Double EWMA Control Scheme" *IEEE Transactions on Semiconductor Man-ufacturing*, 18, 1, pp. 104-111.

20.   Tsung, F., and Shi, J.J. (1999). \Integrated Design of Run to Run PID Controllers and SPC Monitoring For Process Disturbance Rejection," *IIE Transactions*, 31(6), pp. 517-527.

21.   Tsung, F., and Tsui, K.-L. (2003). \A mean-shift pattern study on the integration of SPC and APC for process monitoring", *IIE Transactions*, 35, pp. 231-242.

22.   Tsung, F., Wu, H., and Nair, V., (1998), \On the E±ciency and Robustness of Discrete Proportional-Integral Control Schemes", *Technometrics*, 40, 3, 214-222.

23.   Tucker, W.T., Faltin, F.W., and Vander Wiel, S.A. (1993). \ASPC: an ellaboration", Tech-nometrics, 35, 4, pp.363-375.

24.   Vander Wiel S.A., Tucker, W.T., Faltin, F.W., and Doganaksoy, N. (1992). \Algorithmic Statistical Process Control: Concepts and an Application", Technometrics, 34, 3, pp. 286-297.