THE PRIMAL STATE ADAPTIVE CONTROL CHART

Bruce Hoadley

and

Bill Huston

Bell Communications Research

The Quality Measurement Plan (QMP) is a new kind of control chart based on a hierarchical Bayes model. QMP has been successfully implemented at AT&T and Bellcore. Two features of QMP are potential limitations: (i) all past observations in an arbitrary moving window are treated equally and (ii) conditional on the hyper-parameters (process average and variance), the serial correlation is zero. The Primal State model avoids these limitations. The basic idea is that for each period, the true defect rate does not change with probability (1-P); but with probability P, the defect changes to a random Primal State. We present a recursive adaptive filter for this model and make comparisons to QMP. A fascinating result is that for a very large set of real control chart data, QMP forecasts as well as the Primal State model. This indicates that QMP is good enough for practical purposes.

1. Introduction and summary.

The Quality Measurement Plan (QMP) (Hoadley, 1981) was implemented throughout AT&T Technologies (formerly called Western Electric) in 1980 and Bell Communications Research in 1984. QMP is a statistical method for analyzing discrete time series of quality audit data consisting of defects and their

---

Key words and phrases. Control chart, Empirical Bayes, Adaptive Filtering Time Series, Quality Audits, Acceptance Sampling, QMP, MIL-STD-105D, Statistical Process Control, Hierarchical Bayes, Recursive Estimation.

AMS 1980 Subject classifications: 62NXX Engineering Statistics, 62N10 Quality Control.

expectancy (expected number of defects given standard quality). The process model for QMP is Bayes empirical Bayes or hierarchical Bayes; i.e., the time series of true quality parameters is i.i.d. with unknown process average and variance, which have a known joint prior distribution. The graphics for QMP form a very effective control chart, which provides point and interval estimates of quality as well as traditional tests of the "In Control" hypothesis. In addition, one is provided with graphical displays of estimated process average, process variance, and sampling variance. The estimate of current quality is a shrinkage of the current sample estimate towards the estimated process average. The proportion shrinkage is an adaptive estimate of the ratio of [sampling variance] and [(sampling variance) + (process variance)].

Bell Communications Research (Bellcore) has also implemented a version of QMP (Hoadley, 1984) for both their factory quality audits and their BELLCORE-STD sampling plans (Brush, 1984 and Guyton, 1985). These sampling plans invoke QMP for computing the Process Control Factor (PCF), which is the probability that the next lot will be substandard. The sample sizes are proportional to $\sqrt{PCF}$. QMP is also used to compute the Allowance Number (AN) for lot or system Rejection.

QMP works well for most applications, but has some alleged deficiencies. One is the arbitrary fixed window length for the data. Another is the assumption that the true defect rates for successive lots are uncorrelated, given the process distribution. This leads to equal treatment of past data. These difficulties seem to be most noticeable for data with small expectancy such as 0.15. Such examples arise when applying the BELLCORE-STD sampling plans. These properties of QMP have motivated further research in this area. An extension of QMP is the Quality Evaluation Plan (QEP) (Phadke, 1982). The QEP model features: (i) a square root transformation to stabilize the variance and obtain an approximate Normal distribution and; (ii) a random walk model with unknown drift variance for the process average. This model results in an adaptive recursive Kalman filter for which the drift variance and the process variance are estimated adaptively. QEP solves some of the deficiencies of QMP, but there is a drawback. The model breaks down for small expectancy such as 0.15. In this case the square root transformation does not stabilize the variance and the sample index is too discrete to be considered normally distributed.

The Primal State model provides an alternative approach to adaptively filtering time series of defect rates, and does not have the above deficiencies. The basic idea is that the defect rate at time t is equal to the defect rate at time t-1 with probability (1-P). But with probability, P, the defect rate changes and is chosen at random from a distribution called the Primal State. The change probability and the Primal State distribution are unknown. So QMP is a special case with P=1.

In this paper we derive an approximate recursive adaptive filter for the Primal State model. We then compare this filter with QMP with particular focus on the above deficiencies. We offer encouraging evidence that the Primal State filter provides a robust estimation oriented control chart, which allows for sudden changes in quality, drifts in quality, and QMP like stationary process variation. However, from a forecast point of view, Primal State offers no advantage over QMP when applied to a large real data set (See Section 3.3). In any case, more research is needed, and should include: (i) a Monte Carlo study of estimation error for various process models, and (ii) a Monte Carlo study (like Roberts, 1966) of expected waiting time to detect a shift in the process.

In Section 3.1, we explore the issue of lot rejection for lot-by-lot acceptance sampling with a small expectancy of 0.15. Our rule is to reject the lot if the posterior probability of substandard quality exceeds 0.85. We find that the rejection decision depends heavily on both the process model and the prior distribution of the process parameters.

2. The Primal State.

For time period (or lot or system) t, t=1,2,..., let

n_t = Sample size

x_t = Defects observed in the sample

λ_t = True defect rate in defects per unit.

We assume that the sampling distribution of x_t given λ_t is: Poisson $(n_t\lambda_t)$. We transform to a scale using λ_0 = standard (or objective) value of λ_t. Let

$e_t = E(x_t|\lambda_t = \lambda_0 = n_t\lambda_0)$ = Expectancy

$I_t = x_t/e_t$ = Sample quality index

$\Theta_t = \lambda_t/\lambda_0$ = True quality index.

The expectancy, e_t, plays an important role. Notice that

$$x_t|\Theta_t \text{ is Poisson}(e_t\Theta_t);$$

so, expectancy is the carrier of the sample size information, when working on an index scale.

The Primal State process model is:

$$\Theta_t = \begin{cases} \Theta_{t-1} & \text{with probability } 1-P, \\ \text{Primal State}, & \text{with probability } P, \end{cases}$$

where the Primal State is an independent Gamma random variable with unknown mean and variance, Θ and γ^2. We assume that $(\Theta_0,P,\Theta,\gamma^2)$ have a joint distribution. Note that if P=1, then this model reduces to the QMP model. Also note that given (P,Θ,γ^2), the serial correlation of Θ_t and Θ_{t-1} is (1-P). The motivation for the Primal State model is to allow for sudden changes in quality, serial correlation, drifts, and QMP like stationary process variation.

2.1 Notation and Preliminaries.

We use two notational conventions that could cause confusion. The symbol e is used for expectancy and exp is used as the base of the natural logarithm. The symbol E is used as the expectation operator whenever it is followed by a parentheses or bracket; otherwise, E is used as a variable.

2.1.1 Gamma Distribution.

Let $g(\Theta|x,e) = \Theta^{x-1}(e)^x \exp\{-e\Theta\}/\Gamma(x)$ denote the Gamma density with $\Theta \geq 0$, shape parameter x, and scale parameter 1/e. The associated mean, variance and second moment are x/e, x/e^2 and $x(x+1)/e^2$. If we let n and v denote the mean and variance respectively, then $x = n^2/v$ and $e = n/v$. Also let $S(y|n,v)$ denote $\Pr\{\Theta \geq y\}$, where Θ is the Gamma random variable.

2.1.2 Negative Binomial Distribution.

If $x|\Theta$ is Poisson $(e\Theta)$ and Θ has density $g(\Theta|X,E)$, then $x|X,E$ is Negative Binomial with probability mass function

$$n(x|e,X,E) = \frac{\Gamma(X + x)}{x!\Gamma(x)} [\frac{e}{E + e}]^x [\frac{E}{E + e}]^X .$$

Two useful properties of the Negative Binomial random variable, x, are:

(N1) $$V(x) = e^2X/E^2 + eX/E$$

(N2) $$V[x(x-1)] = v(eX/E,1/X),$$

where

$$v(z,y) = 2z^2(1 + y)\{1 + 2z(1 + 2y) + z^2y(2 + 3y)\}.$$

These results can be derived from the following facts. Let $M_\nu = x(x-1)\ldots(x-\nu+1)$. In Haight (1967) it is shown that $E(M_\nu) = e^\nu E(\Theta^\nu)$. From Kendall (1958), p.64, it follows that

$$V(x) = E(M_2) + E(M_1) - [E(M_1)]^2$$

$$V[x(x-1)] = 2E(M_2) + 4E(M_3) + E(M_4) - [E(M_2)]^2.$$

And from Wilks (1962), p.171, we have $E(\Theta) = X/E$, $E(\Theta^2) = [X/E]^2(1+1/X)$,

$E(\Theta^3) = [X/E]^3(1+1/X)(1+2X)$, $E(\Theta^4) = [X/E]^4(1+1/X)(1+2/X)(1+3/X)$.

2.1.3 Gamma - Conditional Gamma Bayesian Model.

In Appendices D and E of Hoadley (1984), we derive the following result. Let R_1,ω be a statistic, parameter for which $R_1|\omega \sim$ Gamma with mean = $1/\omega$ and shape parameter a_1. Let ω be distributed as a Conditional Gamma $\phi|(\phi \leq 1)$, where $\phi\sim$ Gamma with mean = $1/R_0$ and shape parameter a_0. Then $E(\omega) = 1/[R_0F(a_0,R_0)]$ and $E(\omega|R_1) = 1/[RF(a,R)]$, where $a = a_0 + a_1$, and $R = (a_0R_0 + a_1R_1)/a$, $F(a,R) = G_a(aR)/G_{a+1}(aR)$ and $G_a(y) = \int_0^y g(x|a,1)dx$. Computational formulas for F(a,R) are given in Hoadley (1984), Appendix D.

2.1.4 Beta density.

Let $b(P|A,B) = \Gamma(A+B)P^{A-1}(1-P)^{B-1}/\Gamma(A)\Gamma(B)$ denote the Beta density with $0 \leq P \leq 1$, and parameters A, B. The associated mean, variance and second moments are $A/(A+B)$, $AB/(A+B)^2(A+B+1)$, and $[A/(A+B)]^2[1+B/A(A+B+1)]$. If P has the above Beta distribution and $m = E(P)$, $r = [E(P) - E(P^2)]/V(P)$, then $A = rm$, $B = r(1-m)$.

2.1.5 Notation for manipulating conditional probabilities.

Let $p(\Theta_t|.)$, $p(P|.)$, denote posterior densities of parameters given events or data. When conditioning, we use the symbol 't' to mean 'data through period t'. Some events of interest are C_t = { process changed at period t} and $\overline{C}_t$ = {process remained the same at period t}. So for example, $p(\Theta_t|t,\overline{C}_t)$ is the posterior density of Θ_t given all the data through period t and the process did not change at period t. Also for example, $Pr\{x_t|t-1,\overline{C}_t\}$ is the probability of observing x_t given all the data through period t-1 and the process does not change at period t.

2.2 Heuristic recursive adaptive solution to the primal state model.

The exact solution to the Primal State model is intractable and non-recursive; so in this Section, we present a heuristic, recursive, adaptive algorithm for approximating the results of interest. The heuristic argument itself will be presented in Section (2.3).

Input Data:

$$\{x_t, e_t\} \quad t=1,2,\ldots$$

Fixed Parameters:

δ_1, δ_2: Smoothing constants used to estimate the Primal mean and variance

Θ_0, v_0: Prior estimates of the Primal mean and variance

b: Bad quality level used for the probability that future quality will be bad.

Initial Values of Sufficient Statistics:

$$\hat{I}_0,\ Q1_0,\ \hat{G}_0,\ Q2_0,\ \hat{\Theta}_0,\ V_0,\ A_0,\ B_0,\ F_0,\ L_0.$$

These statistics are defined in the recursive algorithm later in this Section.

Derived Initial Values:

$$\hat{P}_0 = A_0/(A_0 + B_0),\ X1_0 = \hat{\Theta}_0^2/V_0,\ E1_0 = \hat{\Theta}_0/V_0.$$

Recursive Algorithm:

For each period t, after the input data (x_t, e_t) are observed, the following 42 step recursive algorithm gives the results of interest, which are $Pr\{C_t|t\}$, $E(P|t)$, $V(P|t)$, $E(\Theta_t|t)$, $V(\Theta_t|t)$, and $Pr\{\Theta_t > 1|t\}$. These are computed in steps (S22), (S23), (S25), (S33), (S34), and (S37) respectively.

(S1) $I_t = x_t/e_t$ = Sample index

(S2) $K_t = |I_t - F_{t-1}|/[\Theta_0/e_t]^{1/2}$ = Relative forecast error

(S3) $L_t = L_{t-1} + K_t$ = Cumulative relative forecast error

(S4) $M_t = L_t/t$ = Average relative forecast error

(S5) $G_t = x_t(x_t - 1)/e_t^2$ = Estimate of the Primal second moment for period t

(S6) $q1_t = v_0 + \Theta_0/e_t \approx V(I_t|\Theta = \Theta_0, \gamma^2 = v_0)$

(S7) $q2_t = v(e_t\Theta_0, v_0/\Theta_0^2)/e_t^4 \approx V(G_t|\Theta = \Theta_0, \gamma^2 = v_0)$

(S8) $W1_t = q1_t/(q1_t + Q1_{t-1} + \delta_1)$ = Weight on $\hat{I}_{t-1}$ in the smoothing of I_t

(S9) $W2_t = q2_t/(q2_t + Q2_{t-1} + \delta_2)$ = Weight on $\hat{G}_{t-1}$ in the smoothing of G_t

(S10) $Q1_t = (1-W1_t)q1_t \approx V(\Theta|t)$

(S11) $Q2_t = (1-W2_t)q2_t \approx V(\Theta^2 + \gamma^2|t)$

(S12) $\hat{I}_t = (W1_t)\hat{I}_{t-1} + (1-W1_t)I_t \approx E(\Theta|t)$ = Smoothing of I_t

(S13) $\hat{G}_t = (W2_t)\hat{G}_{t-1} + (1 - W2_t)G_t$ = Smoothing of G_t

(S14) $a_t = [v_0 + \Theta_0^2]^2/Q2_t$

(S15) $R_t = \hat{G}_t/\hat{I}_t^2$

(S16) $v_t = (\hat{G}_t)F(a_t,R_t) - \hat{I}_t^2$ = Estimate of Primal variance

(S17) $VO_t = v_t + Q1_t$ = Adjusted estimate of Primal variance

(S18) $XO_t = \hat{I}_t^2/VO_t$ = Shape parameter of the estimated Primal Gamma distribution

(S19) $EO_t = \hat{I}_t/VO_t$ = 1/(Scale parameter) for above distribution

(S20) $f_t = n(x_t|e_t,XO_t,EO_t)$ = Likelihood of change to the Primal State

(S21) $g_t = n(x_t|e_t,X1_{t-1},E1_{t-1})$ = Likelihood of no change

(S22) $$P_t = \frac{A_{t-1}f_t}{A_{t-1}f_t + B_{t-1}g_t} \approx \Pr\{C_t|t\}$$

(S23) $$\hat{P}_t = \frac{A_{t-1} + P_t}{A_{t-1} + B_{t-1} + 1} \approx E(P|t)$$

(S24) $$s_t = P_t\left[\frac{A_{t-1} + 1}{A_{t-1} + B_{t-1} + 1}\right]^2 \left[1 + \frac{B_{t-1}}{(A_{t-1} + 1)(A_{t-1} + B_{t-1} + 2)}\right]$$

$$+ [1 - P_t]\left[\frac{A_{t-1}}{A_{t-1} + B_{t-1} + 1}\right]^2\left[1 + \frac{(B_{t-1}+1)}{A_{t-1}(A_{t-1} + B_{t-1} + 2)}\right]$$

$$\approx E(P^2|t)$$

(S25) $u_t = s_t - \hat{P}_t^2 \approx V(P|t)$

(S26) $$r_t = \frac{\hat{P}_t - s_t}{u_t}$$ = Variable used to compute A_t and B_t

(S27) $A_t = r_t\hat{P}_t$

(S28) $B_t = r_t(1 - \hat{P}_t)$

(S29) $X2_t = X0_t + x_t$ = Shape parameter for posterior distribution of Θ_t given change to Primal State

(S30) $E2_t = E0_t + e_t$ = 1/(scale parameter) for above posterior

(S31) $X3_t = X1_{t-1} + x_t$ = Shape parameter for posterior distribution of Θ_t, given no change

(S32) $E3_t = E1_{t-1} + e_t$ = 1/(Scale parameter) for above posterior

(S33) $\hat{\Theta}_t = P_t[\frac{X2_t}{E2_t}] + (1-P_t)[\frac{X3_t}{E3_t}] \approx E(\Theta_t|t)$

(S34) $V_t = P_t[\frac{X2_t(X2_t + 1)}{(E2_t)^2}] + (1-P_t)[\frac{X3_t(X3_t + 1)}{(E3_t)^2}] - \Theta_t^2 \approx V(\Theta_t|t)$

(S35) $X1_t = \hat{\Theta}_t^2/V_t$ = Shape parameter for the posterior distribution of Θ_t

(S36) $E1_t = \hat{\Theta}_t/V_t$ = 1/(Scale parameter) for above posterior

(S37) $S_t = S(1|\hat{\Theta}_1,V_t) \approx Pr\{\Theta_t > 1|t\}$

(S38) $Q5_t\%$: Defined by $S(Q5_t\%|\hat{\Theta}_t,V_t) = 0.95$

(S39) $Q95_t\%$: Defined by $S(Q95_t\%|\hat{\Theta}_t,V_t) = 0.05$

(S40) $F_t = \hat{P}_t\hat{I}_t + (1 - \hat{P}_t)\hat{\Theta}_t \approx E(\Theta_{t+1}|t)$

(S41) $Y_t = \hat{P}_t(V0_t) + (1 - \hat{P}_t)V_t + \hat{P}_t(1 - \hat{P}_t)(\hat{I}_t - \hat{\Theta}_t)^2 \approx V(\Theta_{t+1}|t)$

(S42) $Z_t = S(b|F_t,Y_t) \approx Pr\{\Theta_{t+1} > b|t\}$.

2.3 Explanation of the heuristic.

(S1) - (S4): These steps are used to compute an average relative forecast error for comparison with other algorithms. In (S2), $[\Theta_0/e_t]^{1/2} = SD(I_t|\Theta_t = \Theta_0)$ is used to normalize the error so as to keep the normalization simple and constant for all algorithms.

(S5) - (S19): These steps are designed to estimate the Primal mean and variance. Note that $E[G_t|P,\Theta,\gamma^2] = E[E[G_t|\Theta_t]|P,\Theta,\gamma^2]$. Using results in Haight (1967), this becomes $E[\Theta_t^2|P,\Theta,\gamma^2] = \gamma^2 + \Theta^2$. Similarly, $E[I_t|P,\Theta,\gamma^2] = \Theta$. So I_t and G_t are estimates of the Primal mean and second moment.

The method used in QMP for estimating the process average and variance is to construct approximate Bayes estimates using data in a moving window of periods. As discussed in Section 1, there are disadvantages to the moving window approach. So for the Primal State algorithm, we first smooth I_t and G_t using (S12) and (S13). These smoothers are simple Kalman filters derived, e.g. in the I_t case, from the model: $I_t|\mu_t \sim N(\mu_t, ql_t)$, $\mu_t|\mu_{t-1} \sim N(\mu_{t-1}, \delta_1)$ and $\mu_0 \sim N(\hat{I}_0, Ql_0)$. Of course we do not have Normal distributions here, but the smoothers are reasonable. The formulas in (S6) and (S7) follow from Section (2.1.2) and the assumption that $P = 1$, $\Theta = \Theta_0$, and $\gamma^2 = v_0$.

A simple estimate of the Primal variance is $\hat{G}_t - \hat{I}_t^2$; but this can be negative. So we inflate $\hat{G}_t$ by the inflation factor $F(a_t, R_t)$, which is motivated by the results in Section (2.1.3). The results are applied with $R_1 = \hat{G}_t/\hat{I}_t^2$, $\omega = \Theta^2/(\Theta^2 + \gamma^2)$, $a_1 = a_t$, $a_0 \to 0$, $R_0 \to 0$. Then the Bayes estimate of $\Theta^2/(\Theta^2 + \gamma^2)$is $\hat{I}_t^2/F \cdot \hat{G}_t^{\,2}$, which motivates (S16). In the theory of Section (2.1.3), $a_1 = E^2(R_1|\omega)/V(R_1|\omega)$. Assuming $\hat{I}_t$ is constant, this becomes $a_1 \approx E^2(\hat{G}_t|\omega)/V(\hat{G}_t|\omega)$. Then we assume $E\,(\hat{G}_t|\omega) \approx v_0 + \Theta_0^2$, $V(\hat{G}_t|\omega) \approx Q2_t$. Of course, for the Kalman filter, $Q2_t$ is approximately $V(\Theta^2 + \gamma^2|t)$. Here, we are using the fact that posterior variances often approximate sampling variances. In any case, we are only trying to inflate $\hat{G}_t$ in a reasonable way.

The motivation for (S17) is to account for the additional uncertainty in the Primal State due to estimation of the Primal mean. Formulas (S18) and (S19) follow from Section (2.1.1).

(S20) - (S22): Assume that approximately,

$$\text{(A1)}.\ p(P|t) = b(P|A_t, B_t).$$

Note that

$$P_t = \Pr\{C_t|t\} = \Pr\{C_t|t-1, x_t\}$$

$$\text{(1)} \qquad = \frac{\Pr\{C_t|t-1\}\Pr\{x_t|t-1,C_t\}}{\Pr\{C_t|t-1\}\Pr\{x_t|t-1,C_t\} + \Pr\{\overline{C}_t|t-1\}\Pr\{x_t|t-1,\overline{C}_t\}}$$

By (A1), $\Pr\{C_t|t-1\} = A_{t-1}/(A_{t-1}+B_{t-1})$, $\Pr\{\overline{C}_t|t-1\} = B_{t-1}/(A_{t-1}+B_{t-1})$. By the definition of the Negative Binomial distribution, $\Pr\{x_t|t-1,C_t\} = f_t$, and $\Pr\{x_t|t-1,\overline{C}_t\} = g_t$. Formula (S22) follows by plugging these into (1).

(S23) - (S29): Note that

$$(2) \qquad p(P|t) = \Pr\{C_t|t\}p(P|t,C_t) + \Pr\{\overline{C}_t|t\}p(P|t,\overline{C}_t).$$

And

$$p(P|t,C_t) = p(P|t-1,C_t) \propto p(P|t-1)\Pr\{C_t|t-1,P\}.$$

By (A1) this is $\propto P^{A_{t-1}-1}(1-P)^{B_{t-1}-1}P$; hence, $p(P|t,C_t) = b(P|A_{t-1}+1, B_{t-1})$ and similarly $p(P|t,\overline{C}_t) = b(P|A_{t-1},B_{t-1}+1)$. Plug these into (2) and formulas (S23) and (S24) follow. Formulas (S25) - (S28) follow from Section (2.1.4) and (A1).

(S29) - (S39): For this derivation, we assume that the Primal State is known and equal to the estimates derived earlier. We also assume

$$(A2) \qquad p(\Theta_t|t) = g(\Theta_t|X1_t,E1_t).$$

Note that

$$(3) \qquad p(\Theta_t|t) = \Pr\{C_t|t\}p(\Theta_t|t,C_t) + \Pr\{\overline{C}_t|t\}p(\Theta_t|t,\overline{C}_t).$$

The Primal State density is $g(.|X0_t,E0_t)$; so by the simple Poisson-Gamma Bayesian model (see Hoadley, 1984, Appendix B), $p(\Theta_t|t,\overline{C}_t) = g(\Theta_t|X2_t,E2_t)$. Using (A2), a similar argument yields $p(\Theta_t|t,\overline{C}_t) = g(\Theta_t|X3_t,E3_t)$. Plug these into (3) and formulas (S33) and (S34) follow. Formulas (S35) - (S39) follow from Section (2.1.1).

(S40) - (S42): Note that

$$(4) \quad E(\Theta_{t+1}|t) = \Pr\{C_{t+1}|t\}E[\Theta_{t+1}|t,C_{t+1}] + \Pr\{\overline{C}_{t+1}|t\}E[\Theta_{t+1}|t,\overline{C}_{t+1}].$$

Formula (S40) follows.

Let i_{t+1} be the indicator random variable of change to the Primal State at period t+1. Then

$$
\begin{aligned}
V(\Theta_{t+1}|t) &= E[V(\Theta_{t+1}|t,i_{t+1}|t] + V[E(\Theta_{t+1}|t,i_{t+1})|t] \\
&= \hat{P}_t V(\Theta_{t+1}|t,C_{t+1}) + (1-\hat{P}_t)V(\Theta_{t+1}|t,\overline{C}_{t+1}) \\
&\quad + \hat{P}_t E^2(\Theta_{t+1}|t,C_{t+1}) + (1-\hat{P}_t)E^2(\Theta_{t+1}|t,\overline{C}_{t+1}) \\
&\quad - \{\hat{P}_t E(\Theta_{t+1}|t,C_{t+1}) + (1-\hat{P}_t)E(\Theta_{t+1}|t,\overline{C}_{t+1})\}^2.
\end{aligned} \tag{5}
$$

Formula (S41) follows by algebra and observing that

$$V(\Theta_{t+1}|t,C_{t+1}) = VO_t,\ V(\Theta_{t+1}|t,\overline{C}_{t+1}) = V_t, E(\Theta_{t+1}|t,C_{t+1}) = \hat{I}_t,$$

$$E(\Theta_{t+1}|t,\overline{C}_{t+1}) = \hat{\Theta}_t.$$

Finally, (S42) follows by assuming that $\Theta_{t+1}|t$ is approximately Gamma.

3. Examples.

Throughout this Section we use the same fixed parameters and initial sufficient statistics for the Primal State Algorithm. The fixed parameters are:

$$\delta_1 = \delta_2 = 0.01,\ \Theta_0 = 1,\ v_0 = 0.55,\ b = 3.$$

The initial sufficient statistics are:

$$\hat{I}_0=1,\ Q1_0=3.05,\ \hat{G}_0=1.55,\ Q2_0=1,\ \hat{\Theta}_0=1,\ V_0=3.6,\ A_0=1,\ B_0=1,\ F_0=1,\ L_0=0.$$

Also throughout this Section, we use the standard parameters for the implementation of QMP described in Hoadley (1984), which are

$$\Theta_0=1,\ v_0=2.5,\ \gamma_{mo}^2=0,\ \gamma_o^2=0.55,\ \gamma_{max}^2=2.2.$$

For the application of QMP in the BELLCORE-STD Sampling plans, the window length (T) depends on the expectancies. The window lengths for examples 3.1 and 3.2 are 16 and 6 respectively.

3.1 Small expectancy, flurry of defects.

QMP and MIL-STD-105D (Duncan, 1974, p.209) behave differently with respect to rejecting lots when the expectancy is 0.15. The purpose of this Section is to explore this whole issue by comparing the behavior of the Primal State model vs. QMP and MIL-STD-105D for a specially designed example.

The input data for this example is $e_t \equiv 0.15$, t=1,...,43; $x_t \equiv 0$, t=1,...,17,32,...,43; and for t=18,...,31, the defect pattern is:

t	18	19	20	21	22	23	24	25	26	27	28	29	30	31
x_t	1	0	2	0	1	0	3	0	1	0	2	0	0	1

Figure 1 is a plot of $\Pr\{\Theta_t > 1|t\}$ for QMP vs. Primal State.

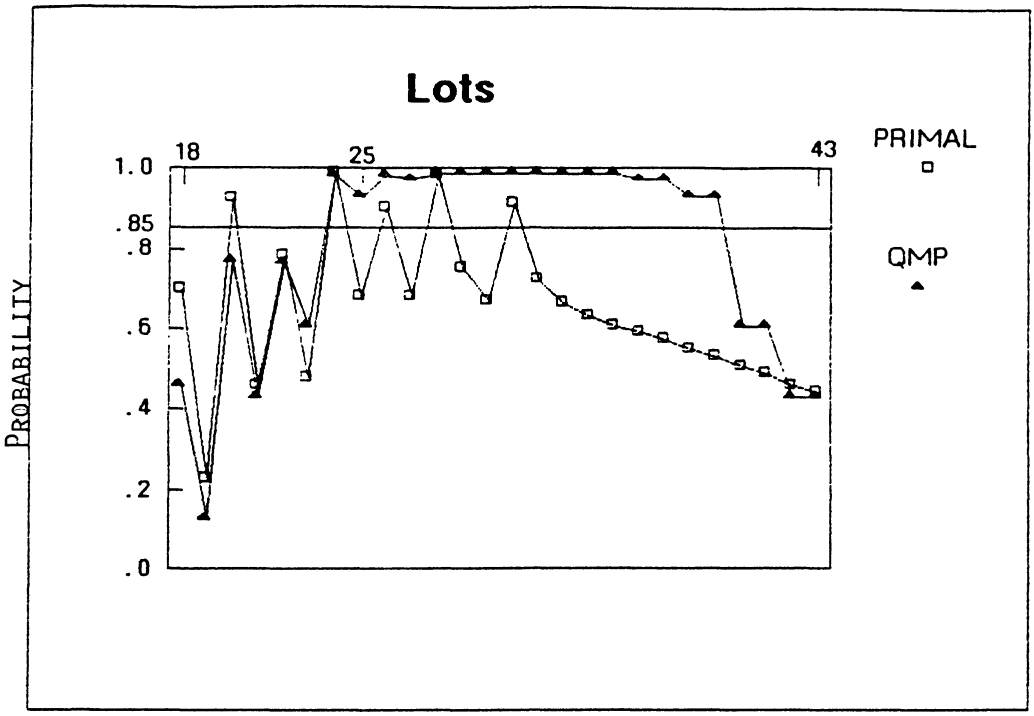

Figure 1. Posterior probability of being substandard for QMP vs. Primal State.

For the Normal State of MIL-STD-105D, the Allowance Number is zero; i.e., a lot with one defect is rejected. The producer's risk is 0.14. For BELLCORE-STD sampling plans, the rejection criterion is: $Pr\{\Theta_t > 1|t\} > 0.85$; so that the Bayes producer's risk is less than 0.15. A comparison of the Accept(A)/ Reject(R) decisions for the various plans is shown below:

Lot (t)	18	19	20	21	22	23	24	25	26	27	28	29	30	31
MIL-STD-105D	A	A	R	A	R	A	R	A	R	A	R	A	A	R
QMP	A	A	A	A	A	A	R	R	R	R	R	R	R	R
PRIMAL STATE	A	A	R	A	A	A	R	A	R	A	R	A	A	R

For lot 18 (the first defect), MIL-STD-105D Accepts because all of the preceding zero defect lots put the plan into the Reduced State. QMP and Primal

State also Accept with $Pr\{\Theta_t > 1|t\}$ being 0.46 and 0.70 respectively. For all lots except lot 22, Primal State agrees with MIL-STD-105D. For lot 22, $Pr\{\Theta_t > 1|t\} = 0.78$ for Primal State - not quite large enough for Rejection. However, the behavior of QMP is completely different. QMP Accepts through lot 23 and then Rejects the rest - even when there are zero defects.

To see why this is true, consider lot 25, which has zero defects, but was preceded by a flurry of defects. For lot 25, a comparison of QMP vs. Primal State reveals:

	$E(\Theta_t\|t)$	$SD(\Theta_t\|t)$	$Pr\{\Theta_t>1\|t\}$	Decision
QMP	2.56	.82	.93	R
Primal	2.20	1.97	.68	A

The difference in decision is mostly due to the different posterior standard deviations. These are in turn mostly due to the difference between the QMP process variance (.73) and the Primal State variance (3.37). These are different because the simple Kalman filter used for smoothing $\{G_t\}$ is initialized with prior information that is much more diffuse than the comparable QMP prior information on process variance. The idea is that the Primal State distribution describes what happens when the process changes from a state of statistical control to something else, which could be dramatically different. The QMP process variance models lot to lot variation which is typically small. The Primal variance is intended to model dramatic change in the process, and is therefore more uncertain.

We have shown that for small expectancy the Accept/Reject decision for a lot depends heavily on both the process model and the prior distribution of the process parameters. This serves as a warning to exercise great care in the model specification. The standard solution to this Accept/Reject decision problem is MIL-STD-105D. This solution is approximately consistent with the Primal State solution presented herein. However, if the QMP model or the Primal State model with a different initialization is more realistic, then MIL-STD-105D is not the solution. Implicit in MIL-STD-105D is some model; and we say, let it be revealed.

3.2 An example with real data.

Preliminary applications of the Primal State algorithm to real data are very encouraging. The kind of example for which the Primal State algorithm works very well (in a forecast error sense) is shown in Figure 2. For this example, the expectancies range from 1.0 to 3.6. The data for periods 1 through 16 are a copy of the data for periods 17 through 32. The solid and dashed curves are $E(\Theta_t|t)$ (called the filter) for QMP and Primal State respectively. The solid and dashed inverted 'T' symbols are the fifth posterior percentiles ($Q5_t\%$) for QMP and Primal State respectively.

Notice that for Primal State, $E(\Theta_t|t)$ is quicker to change and quicker to level off - a pattern which is consistent with the underlying model. The fifth posterior percentiles are quite different for a period like 25, where there appears to be a drift. The ratio of the average absolute forecast errors (averaged over periods 17 through 32) for Primal State vs. QMP, is .84.

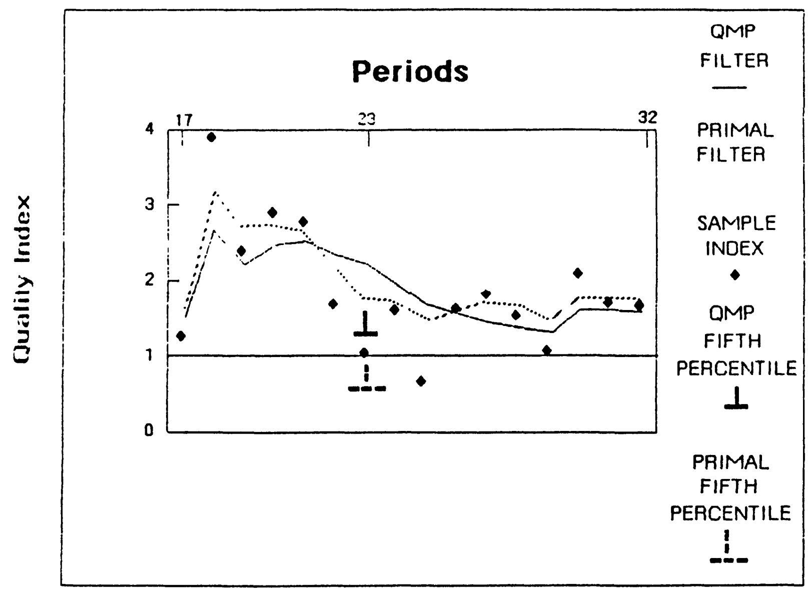

Figure 2. QMP vs. Primal State filters for some real data.

3.3 Forecast error comparison.

In this Section we compare the forecast errors of Primal State vs. QMP (with a 6 period window). Our data base is 1,467 real time series with 16 periods each. Most of the expectancies are between .2 and 10 with an average of 3.8. For each time series, the Primal State Average Relative Forecast Error (ARFE) is defined in (S4) of the recursive algorithm in Section 2.2. For QMP, the ARFE is defined similarly using the process average as the forecast. The arithmetic means of the two sets of 1,467 ARFE's are .8932 and .8914 for Primal State and QMP respectively - a ratio of 1.0020 - signifying no important difference. This suggests that QMP, which is conceptually very simple, is quite adequate for most applications.

REFERENCES

Brush, G.G., Guyton, D.A., Hoadley, B., Huston, W.B., and Senior, R.A. (1984). BELL-STD-100 An Inspection Resource Allocation Plan. Transactions of IEEE Communications Society, November 26-29, 1984 1119-1124, Atlanta.

Duncan, A.J. (1974). Quality Control and Industrial Statistics. Fourth Edition, Richard D. Irwin, Inc.

Guyton, D.A. and Hoadley, B. (1985). BELLCORE-STD-200 System Reliability Test Sampling Plan. 1985 Proceedings Annual Relatibility and Maintainability Symposium, 426-431.

Haight, F. (1967). Handbook of the Poisson Distribution. Wiley, New York.

Hoadley, B. (1981). The Quality Measurement Plan (QMP). Bell System Technical Journal **60** 215-273.

Hoadley, B. (1984). QMP Theory and Algorithms. Bell Communications Research Technical Memorandum TM-TSY-000238, October 26, 1984 (available for public release from author).

Kendall, M.G. and Stuart, A. (1958). The Advanced Theory of Statistics, Volume 1, Distribution Theory. Charles Griffin & Co. Ltd., London.

Phadke, M.S. (1982). Quality Evaluation Plan Using Adaptive Kalman Filtering. Bell System Technical Journal **61**, 2081-2107.

Roberts, S.W. (1966). A comparison of some control chart procedures. _Technometrics_ 8 411-430.

Wilks, S.S. (1962). _Mathematical Statistics_. Wiley, New York.