

# The GLRT for statistical process control of autocorrelated processes

DANIEL W. APLEY<sup>1</sup> and JIANJUN SHI<sup>2</sup>

<sup>1</sup>*Department of Industrial Engineering, Texas A & M University, College Station, TX 77843-3131, USA*  
E-mail: [apley@tamu.edu](mailto:apley@tamu.edu)

<sup>2</sup>*Department of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, MI 48109-2117, USA*  
E-mail: [shihang@engin.umich.edu](mailto:shihang@engin.umich.edu)

Received November 1996 and accepted December 1998

---

This paper presents an on-line Statistical Process Control (SPC) technique, based on a Generalized Likelihood Ratio Test (GLRT), for detecting and estimating mean shifts in autocorrelated processes that follow a normally distributed Autoregressive Integrated Moving Average (ARIMA) model. The GLRT is applied to the uncorrelated residuals of the appropriate time-series model. The performance of the GLRT is compared to two other commonly applied residual-based tests – a Shewhart individuals chart and a CUSUM test. A wide range of ARIMA models are considered, with the conclusion that the best residual-based test to use depends on the particular ARIMA model used to describe the autocorrelation. For many models, the GLRT performance is far superior to either a CUSUM or Shewhart test, while for others the difference is negligible or the CUSUM test performs slightly better. Simple, intuitive guidelines are provided for determining which residual-based test to use. Additional advantages of the GLRT are that it directly provides estimates of the magnitude and time of occurrence of the mean shift, and can be used to distinguish different types of faults, e.g., a sustained mean shift versus a temporary spike.

## 1. Introduction

With increasing automation, continuous flow and discrete parts manufacturing processes are being measured at higher and higher rates. It is no longer uncommon to find 100% inspection of discrete processes. As the process is measured at higher rates, the data, to be used for quality control and/or process control purposes, is more likely to be autocorrelated. It is well known that the run length properties of CUSUM tests are strongly affected by data correlation, and the in control Average Run Length (ARL) can be much shorter than claimed if the autocorrelation is positive [1]. Vasilopoulos and Stamboulis [2] have shown that the same is true for  $\bar{X}$  charts.

One obvious solution is to modify the control limits, based on the autocorrelation structure, in order to achieve the desired in control ARL. Yashchin [3] provides a method for approximating the ARL in the presence of autocorrelation and recommends applying a CUSUM to the original data if the autocorrelation is not too strong. One could also monitor the batch-means of the process, which are less affected by autocorrelation [4,5]. An alternative, increasingly popular, scheme is to model the autocorrelated data using Autoregressive Integrated Moving-Average (ARIMA) time-series models. Such

models can effectively describe a wide variety of processes and are simple enough in structure to allow convenient analysis [6]. If the model is accurate, the model residuals (specifically, the minimum-mean-square-error, one-step-ahead prediction errors) are approximately uncorrelated, and any of the standard SPC techniques can be applied to the residuals with well understood in control run length properties [7–11].

When the process experiences a mean shift, however, the run length properties are less well understood. Since the residuals are a filtered version of the original process, the resulting mean shift in the residuals will not have the same form as the mean shift in the original process. Rather, it will experience some dynamics, which depend on the ARIMA model, before settling down to a steady-state value. We refer to this time-varying mean of the residuals as the “fault signature,” since it is a pattern, or signature, left in the residuals by a mean shift fault in the original data. The particular form of the fault signature depends heavily on the ARIMA model that describes the original process. Hu and Roan [12] have suggested visual inspection of these patterns for distinguishing between step mean shifts and spikes. Various authors have used theoretical derivation, Markov chain approximation, Monte Carlo simulation, or combinations of the three techniques to investigate the in-control and out-of-

control run length properties of standard monitoring schemes such as Shewhart, EWMA, and CUSUM tests, applied to the residuals [13–17]. One common conclusion is that the signaling performance of standard tests is often disappointingly poor, due mainly to what has been referred to as “forecast recovery” [15]. Forecast recovery means that any significant effect (i.e., in terms of the fault signature) the fault has on the residuals is short lived, and the lasting effects are minor.

For many ARIMA models that suffer from forecast recovery, however, the fault signature has considerable transient dynamics. Neither Shewhart, EWMA, nor CUSUM tests make use of this valuable information. In contrast, a Generalized Likelihood Ratio Test (GLRT) can automatically take this information into account. Because of this desirable feature, the GLRT has been used for years in failure detection in dynamic systems, such as aircraft [18], using state-space system models and Kalman filters for generating residuals. However, it was not until recently that this potential was exploited for the purpose of SPC for autocorrelated data. Apley and Shi [19] developed a GLRT for SPC of autocorrelated processes described by a particular second order ARMA model and found the GLRT to perform significantly better than a CUSUM test on the residuals. In contrast, for first order IMA processes, van der Wiel [16] compared a GLRT, Shewhart, EWMA, and CUSUM test on the residuals and found the CUSUM test to be superior.

The purpose of this paper is to: (i) elaborate on the GLRT for mean shift detection in autocorrelated processes; (ii) present a run length comparison between the GLRT and the residual-based Shewhart and CUSUM tests for step mean shifts in a variety of ARIMA processes; and (iii) provide a simple means of determining which residual-based test should be most effective for a given ARIMA model. It will be demonstrated that which test is most effective depends predominantly on the particular shape of the fault signature and thus on the ARIMA model used to describe the process. In situations where the GLRT performs significantly better, it is because there are pronounced fault signature dynamics that the GLRT is able to exploit, but that the other tests ignore.

## 2. Time series residuals and the fault signature

### 2.1. The basic model

ARIMA models are a fairly general class that can be used to describe a wide variety of autocorrelated processes [6], and their applicability as SPC tools for autocorrelated processes is now widely known. This section briefly reviews ARIMA models and how to use them to generate uncorrelated residuals and illustrates the fault signatures associated with a variety of ARIMA models.

The general model for a Gaussian ARIMA( $p, d, q$ ) process  $x(t)$  is

$$x(t) = \frac{\Theta(B)}{(1-B)^d \Phi(B)} a(t), \quad (1)$$

where  $a(t)$  is an i.i.d. Gaussian process with mean 0 and variance  $\sigma_a^2$  denoted  $a(t) \sim \text{NID}(0, \sigma_a^2)$ ,  $t$  is a time index,  $B$  is the backshift operator, and  $\Phi(B)$  and  $\Theta(B)$  are polynomials of degree  $p$  and  $q$ , respectively, with roots lying outside the unit circle.  $\Phi(B)$  and  $\Theta(B)$  are referred to as the AR and MA polynomials, respectively, and are parameterized as  $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$  and  $\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ .

It is assumed that the autocorrelated process to be monitored is given by

$$y(t) = x(t) + \mu f(t), \quad (2)$$

where  $f(t)$  is a deterministic unit magnitude mean shift fault (as a function of time) occurring in the original data, and  $\mu$  is the magnitude of the fault. Although other forms of faults (e.g., ramps or temporary spikes) can easily be considered, the major focus of this paper will be step mean shifts of the form

$$f(t) = \begin{cases} 0, & t < \tau, \\ 1, & t \geq \tau, \end{cases} \quad (3)$$

where  $\tau$  is the time of occurrence of the mean shift. “Out of control condition”, “assignable cause”, “special cause”, “unusual event”, etc., are terms commonly used in the literature that refer to situations that result in such faults.

Since the model of (1) is invertible, uncorrelated residuals can be generated by filtering  $y(t)$  with the “inverse model”, obtained by interchanging the numerator and denominator of the ARIMA model. Denoting the residual process by  $e(t)$ , the equation for generating  $e(t)$  is

$$e(t) = \frac{\Phi(B)(1-B)^d}{\Theta(B)} y(t). \quad (4)$$

Substituting (1) into (4) gives

$$e(t) = a(t) + \mu \tilde{f}(t), \quad \text{where} \quad (5)$$

$$\tilde{f}(t) \equiv \frac{\Phi(B)(1-B)^d}{\Theta(B)} f(t), \quad (6)$$

is the fault signature of the fault  $f(t)$  as it manifests itself in the residuals. (5) is the basis for applying SPC techniques to the residuals of an autocorrelated process. Under no fault conditions  $f(t)$ , and thus  $\tilde{f}(t)$ , is identically zero. The residuals are then the original zero-mean i.i.d. Gaussian process  $a(t)$ , and standard SPC techniques designed for uncorrelated data can be applied with well understood in-control run length properties. On the other hand, when a fault does occur the residuals are  $a(t)$  plus the deterministic fault signature. Thus, they are uncorrelated and Gaussian with variance  $\sigma_a^2$  and with a time-

varying mean that is equal to the fault signature (scaled by  $\mu$ ). This follows by taking the expectation of (5), which gives

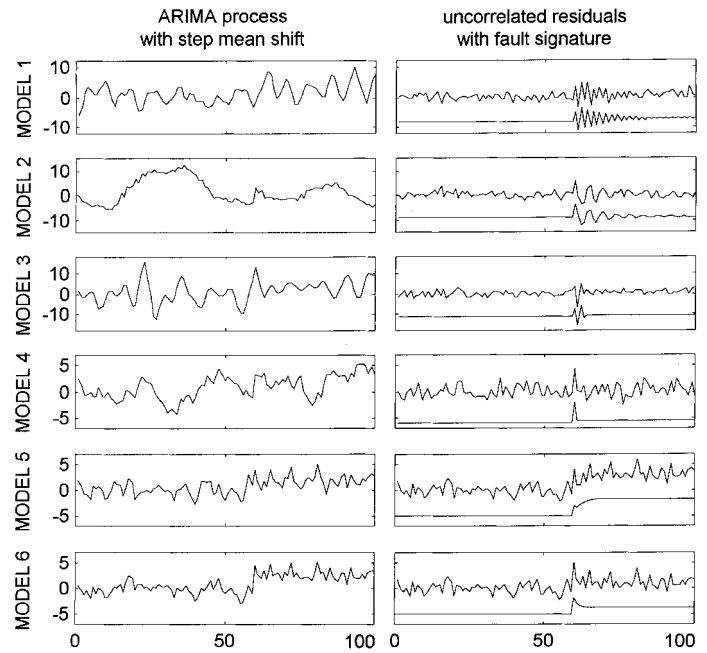
$E[e(t)] = E[a(t) + \mu\tilde{f}(t)] = E[a(t)] + E[\mu\tilde{f}(t)] = \mu\tilde{f}(t)$ . Consequently, if a fault of significant magnitude occurs, it can leave a pronounced signature in the residuals being monitored.

These results, of course, depend on the model being accurately known. Unless otherwise noted, it will be assumed throughout the remainder of the paper that this is the case. A thorough investigation of the effects of modeling inaccuracies is beyond the scope of this paper, but a brief analysis will be presented in Section 4.3.

**2.2. Illustrative examples of fault signatures and the associated ARIMA models**

The particular form of the fault signature  $\tilde{f}(t)$  depends heavily on the ARIMA model. Using standard terminology from linear systems theory,  $\tilde{f}(t)$  (for a step mean shift fault) is the dynamic step response of the inverse model. It is important to note that if a step mean shift occurs in the original process, the resulting mean shift in the residuals will not be a step function. Rather, it will experience some transient dynamics before eventually settling down to a steady-state value. As will be demonstrated shortly, the ability of various residual-based tests to detect mean shifts depends heavily on the form of the fault signature. To illustrate these concepts, the following six examples of different types of ARIMA models, all with different fault signatures, will be used throughout the paper. The behavior of the models is illustrated in Fig. 1, and Table 1 lists the AR polynomial, MA polynomial, and  $d$  for each model. Without loss of generality,  $\sigma_a$  was chosen to be one for all models.

The first column of Fig. 1 shows the original autocorrelated data with a step mean shift occurring at time 60 for each of the six models.  $\mu = 3, 5, 4, 4, 2,$  and  $3,$  respectively, for Models 1 through 6. Note that although specific values of  $\mu$  were chosen for illustrating the processes in Fig. 1, a broad range of values of  $\mu$  will be used in Section 4.2 when analyzing and comparing the various tests. For many of the models it is difficult to discern the mean shift due to the high level of autocorrelation of the data. The second column shows the residuals, generated



**Fig. 1.** Illustration of six ARIMA processes and their corresponding fault signatures for a step mean shift. The parameters of the six models are listed in Table 1. Original autocorrelated ARIMA processes with step mean shifts at timestep 60 are shown in the left column. The right column shows the uncorrelated residuals with the fault signature for each of the six processes. The fault signature itself, which is offset vertically for clarity, is shown below the residuals.

from (4), for the same six processes. The fault signatures themselves are shown directly below the residual processes. Note that the fault signatures have the same scale, except that they have been offset vertically from the residuals for better illustration. The flat portion of the fault signature, before the fault has occurred, is taken to be zero.

The six models were selected not to represent any specific processes, but so that the general shapes of the corresponding fault signatures cover as wide a variety as possible (see Fig. 1). It is reasonable to expect that any of these models could be encountered in practical applications. In certain situations one can argue the validity of a particular ARIMA model based on physical modeling. English and Case [20] argue that continuous flow pro-

**Table 1.** The six ARIMA models used throughout this paper

	$\Phi(B)$	$\Theta(B)$	$d$
Model 1	$1 - 1.13B + 0.64B^2$	$1 + 0.9B$	0
Model 2	1	$1 - 0.31B + 0.81B^2$	1
Model 3	$1 - 2.19B + 2.39B^2 - 1.4B^3 + 0.41B^4$	1	0
Model 4	$1 - 0.9B$	1	0
Model 5	$1 - 0.99B + 0.49B^2$	$1 - 0.7B$	0
Model 6	$1 - 0.8B$	$1 - 0.5B$	0

cesses such as chemical and petroleum production processes are well modeled as a cascaded series of first order linear differential equations. When such continuous processes are sampled, they become  $ARIMA(p, 0, p - 1)$  models, where  $p$  is the number of cascaded first order systems. As an example, consider four cascaded subsystems with individual time constants of 1, 2, 3, and 4 minutes. If such a system is sampled every 0.4 minutes, the resulting discrete time model is  $ARIMA(4,0,3)$  with fault signature given by curve (a) in Fig. 2. This fault signature has oscillatory behavior similar to that of Models 1 through 3.

Although many production processes are far too complicated to model physically, they may still be well modeled as ARIMA processes. In these situations, ARIMA modeling can be thought of as a somewhat generic way to describe the nature of the dynamics, or autocorrelation, of the process [6]. Consequently, a wide variety of ARIMA models may appear in practical examples of real production processes. MacGregor and Harris [21] modeled the dry basis weight in a paper making process as an  $ARMA(1,1)$  process with  $\phi_1 = 0.81$  and  $\theta_1 = 0.51$ , which has the fault signature shown as curve (b) in Fig. 2. This model is nearly identical to Model 6. Hu and Roan [12] found an  $ARIMA(1,0,0)$  model with  $\phi_1 = 0.99$  closely fitted dimensional data in an autobody assembly process. The fault signature for this model is shown as curve (c) in Fig. 2 and is very similar to that of Model 4.

Figure 3(a) shows dimensional data from another autobody assembly process. All data is in units of mm. The high degree of negative autocorrelation at lag 1 is the result of multiple tooling. It was found that an  $ARIMA(4,0,3)$  model closely fits the data with  $\Phi(B) = 1 + 0.62B + 0.32B^2 + 0.26B^3 - 0.31B^4$  and  $\Theta(B) = 1 - 0.17B + 0.59B^2 + 0.26B^3$ . It was known *a priori* that a shim move upstream caused a sudden mean shift (with an

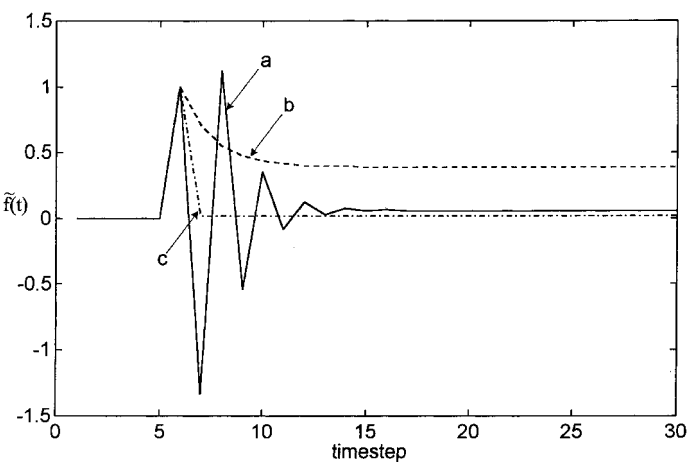


Fig. 2. Fault signatures for ARIMA models of various production processes: (a) chemical production process [20]; (b) dry basis weight in a paper making process [21]; and (c) dimensional data in autobody assembly [12].

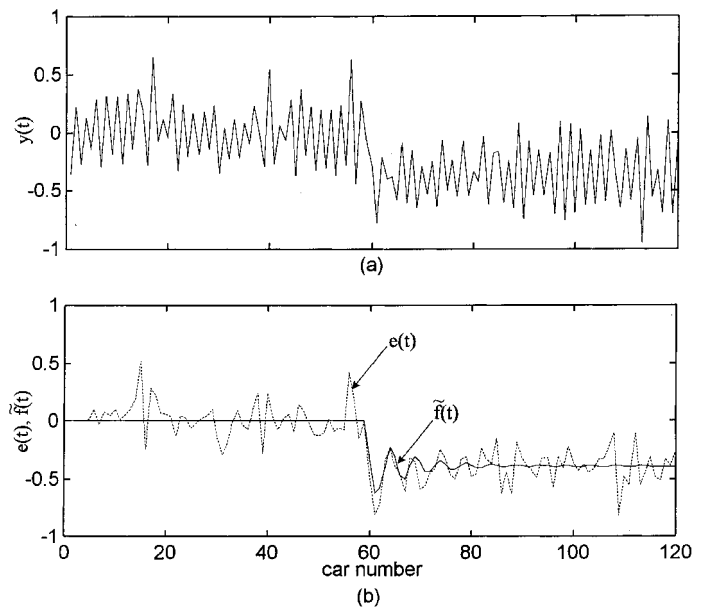


Fig. 3. Dimensional data (mm) in an autobody assembly process with a sudden mean shift at car number 60: (a) raw data  $y(t)$ ; and (b) ARIMA residuals  $e(t)$  and fault signature  $\tilde{f}(t)$ .

estimated magnitude of 0.35 mm) in the data at car number 60. Step mean shifts are common in autobody assembly, the result of new batches of stamped parts being used, as well as the result of shim moves. The large process variation caused by the multiple tooling tends to mask the presence of the mean shift. The ARIMA residuals of the process are shown in Fig. 3(b), along with the fault signature. Not only is the mean shift more evident in the residuals, but also the oscillatory dynamics of the fault signature are clearly present in the residuals. Such oscillatory dynamics, like those of Models 1 through 3, are characteristic of many ARIMA models. It should be noted that for a step mean shift fault, the form of the fault signature uniquely determines the ARIMA model and *vice-versa*.

### 2.3. Fault signature characteristics

For each of Models 1 through 4, the fault signatures settle down to steady-state values that are quite small. This is referred to as forecast recovery. It is intuitively obvious that a CUSUM test applied to these residuals would perform very poorly. For the first three examples, the initial oscillation of the fault signature would have a tendency to reset the CUSUM statistic to zero when the fault signature becomes negative. Moreover, if the mean shift is not detected very soon after its occurrence, it is likely that it will not be detected for a long time, since the steady-state magnitudes of the fault signatures are so small. Thus, there is a clear window of opportunity for detecting the mean shift. A Shewhart test may perform somewhat better than a CUSUM test due to the peaks

during the transient phase of the fault signature, but is still subject to the window of opportunity.

In spite of the low steady-state value, the fault signature for each of the first three models experiences significant dynamics in the transient stage. It was shown in the paragraph following (6) that the fault signature multiplied by  $\mu$  is exactly the mean of the residuals. This follows since  $a(t)$  is assumed to be zero mean and  $\tilde{f}(t)$  is assumed to be deterministic. Consequently, the dynamics of the fault signature are highly visible in the residuals. Neither the Shewhart, nor the CUSUM, test is designed to make use of this valuable information. In contrast, the GLRT presented in the subsequent section does exactly this. It makes explicit use of the fault signature dynamics by, in some sense, testing for the presence of these dynamics in the residuals. As will be demonstrated in Section 4, for Models 1 through 3 the GLRT has drastically improved performance over a Shewhart or CUSUM test on the residuals. For Model 4 the improvement is only slight.

For Models 5 and 6, the situation is different. The steady-state magnitude of the fault signature is considerably larger than for the first four models, and thus one could conjecture that a CUSUM test on the residuals would perform reasonably well. Since the dynamics of the fault signatures for these two models are not pronounced, the GLRT has little extra information to make use of, and the CUSUM test actually outperforms the GLRT. These intuitive statements regarding the performance of the GLRT, Shewhart, and CUSUM tests will be verified quantitatively in Section 4. The main conclusion will be that, given an ARIMA description of an autocorrelated process, the most effective residual-based test can be determined in a simple, intuitive manner based predominantly on the shape of the fault signature.

Given the availability of personal computers, it is a straightforward matter to calculate the fault signature for a particular ARIMA model and a particular fault. This is done recursively using (6), rewritten as  $\Theta(B)\tilde{f}(t) = \Phi(B)(1 - B)^d f(t)$ . Initial conditions are  $\tilde{f}(t) = f(t) = 0$  for  $t < \tau$ , where  $\tau$  is the time of occurrence of the fault.

If  $f(t)$  converges to a steady-state value, the fault signature always converges to a steady-state value, denoted  $\tilde{f}_{ss}$ .  $\tilde{f}_{ss}$  is of particular interest when one is considering the effectiveness of a CUSUM test on the residuals. Applying the final value theorem [22] for discrete-time difference equations to (6) gives

$$\tilde{f}_{ss} = \left. \frac{\Phi(B)(1 - B)^d}{\Theta(B)} \right|_{B=1} f_{ss}, \quad (7)$$

where  $f_{ss}$  is the steady-state magnitude of  $f(t)$ . For a fault of magnitude  $\mu$ , the steady-state magnitude of the mean of the residuals is  $\mu\tilde{f}_{ss}$ . It follows from (7) that for an ARIMA( $p, d, q$ ) model with  $d \geq 1$ , the steady-state magnitude of the fault signature is always 0. This can be observed for Model 2 in Fig. 1.

An intuitive ‘‘proof’’ of the final value theorem used to obtain (7) is as follows. For notational simplicity, assume  $d = 0$ . Expanding (6) gives  $(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)\tilde{f}(t) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)f(t)$ , or, equivalently,  $\tilde{f}(t) - \phi_1 \tilde{f}(t - 1) - \phi_2 \tilde{f}(t - 2) - \dots - \phi_p \tilde{f}(t - p) = f(t) - \theta_1 f(t - 1) - \theta_2 f(t - 2) - \dots - \theta_q f(t - q)$ . Assume that  $\tilde{f}(t)$  does, in fact, reach some constant steady-state value, denoted  $\tilde{f}_{ss}$ . Then, for sufficiently large  $t$ ,  $\tilde{f}(t) \cong \tilde{f}(t - 1) \cong \dots \cong \tilde{f}(t - p) \cong \tilde{f}_{ss}$ , and  $f(t) \cong f(t - 1) \cong \dots \cong f(t - q) \cong f_{ss}$ . Substituting this into the above equation gives  $(1 - \phi_1 - \phi_2 - \dots - \phi_p)\tilde{f}_{ss} = (1 - \theta_1 - \theta_2 - \dots - \theta_q)f_{ss}$ , which is the same as (7) for the special case of  $d = 0$ .

### 3. A GLRT for SPC of autocorrelated processes

Although the GLRT has been extensively used for failure detection in dynamic systems [18,23], it has only recently been considered for SPC purposes. GLRTs suitable for on-line SPC of autocorrelated processes have been proposed for ARMA processes [19] and IMA processes [16]. In this section we present the nearly identical form of the GLRT for ARIMA processes. A brief derivation is provided in the Appendix.

One possibility for implementing the GLRT would be, at time  $t$ , to test for faults occurring at all prior timesteps. The computational expense, however, would be excessive. A more manageable approach taken in this paper is to test only for faults occurring in a window of length  $N$  (a user defined quantity), i.e., between timesteps  $t - N + 1$  and the current time  $t$ . Thus, to implement the test one need only consider the  $N$  most recent residuals. Assuming temporarily that only one type of fault is being tested for, the problem is one of multiple hypotheses testing. The null hypothesis is that no fault has occurred, and each of the  $N$  alternative hypotheses is that the fault occurred at one of the  $N$  timesteps within the window.

At current time  $t$ , the Generalized Likelihood Ratio (GLR) statistic associated with the hypothesis that the fault occurred at time  $t - k + 1$  ( $k = 1, 2, \dots, N$ ) is given by (see the Appendix)

$$T_k(t) = \left( \sigma_a^2 \sum_{i=1}^k \tilde{f}^2(i) \right)^{-1/2} \sum_{i=1}^k e(t - k + i) \tilde{f}(i), \quad (8)$$

where  $\tilde{f}(\bullet)$  is the fault signature for a fault occurring at timestep 1. The interpretation of (8) is straightforward. It is the inner product of the two vectors  $[\tilde{f}(1)\tilde{f}(2)\dots\tilde{f}(k)]$  and  $[e(t - k + 1)e(t - k + 2)\dots e(t)]$ , scaled by the norm of the former and  $\sigma_a$ . As such,  $T_k(t)$  is a measure of the ‘‘correlation’’ between the actual residuals and the signature of a fault assumed to have occurred at time  $t - k + 1$ . The higher the correlation between the residuals and the fault signature, the larger the magnitude of  $T_k(t)$  and the larger the likelihood that the fault occurred

at time  $t - k + 1$ . From Fig. 1 it is clear that if the magnitude of the fault is large enough, the residuals will be highly correlated with the fault signature.

The GLRT then involves comparing

$$G(t) \equiv \max_{k=1, \dots, N} |T_k(t)|, \tag{9}$$

with a user defined threshold  $\gamma$  chosen to provide a desired in control ARL. If  $G(t) < \gamma$  the residuals do not correlate closely with any of the fault signatures, and it is concluded that no fault has occurred. If  $G(t) \geq \gamma$  the residuals are highly correlated with one or more of the fault signatures, and it is concluded that the fault occurred at time  $k^*$ , where  $k^*$  is the value of  $k$  that maximizes (9).

Since, in practice,  $\sigma_a^2$  will not be known, one must use a suitable estimate in (8).  $\sigma_a^2$  can be estimated either at the same time as the ARIMA model or at some later time by estimating the variance of the residuals when it is known that no fault was present. Note that the same factor  $\sigma_a^2$  is present in all of the GLR statistics in (8) and could, therefore, be pulled out of the equation for the test statistic and absorbed into the threshold.

To implement the GLRT, the user must select the test parameters  $N$  and  $\gamma$ . For all examples in this paper  $N = 20$  was chosen. Smaller values of  $N$  (e.g., 5 or 10), with  $\gamma$  modified to provide the same in control ARL, offer a slightly higher probability of detection during the early timesteps. However, use of the larger window length of 20 was warranted by the added security of increasing the probability of detection after a reasonable amount of time. For Models 1 through 4, window lengths larger than 20 had little added benefit, since by timestep 20 the fault signatures decay to very small values. From (8), if one computed the GLR statistic for a fault occurring 40 (for example) timesteps prior to the current time, very little weight would be placed on the most recent residuals.

For a desired in-control ARL, the required  $\gamma$  will depend on the fault signature and  $N$ . The GLR statistics can be put into the form of an  $N$ -dimensional Markov chain that is stationary under no-fault conditions. Since the state transition matrix is primitive, by the same arguments presented in Brooks and Evans [24] for the CUSUM test, the tail of the in-control run length distribution is geometric. Monte Carlo techniques for determining  $\gamma$  can thus be more efficiently implemented by truncating the simulations and estimating the geometric decay constant to approximate the tail. For a discussion on the geometric properties of the tail of the run length distribution, see Woodall [25].

Aside from any potential run length improvements, the GLRT possesses some attractive properties that Shewhart and CUSUM tests do not. The GLRT not only detects faults, but also estimates the fault magnitude and time of occurrence and automatically classifies faults according to type if more than one type is hypothesized. If a signal is sounded, the Maximum Likelihood Estimate

(MLE) of the time of occurrence is  $t - k^* + 1$ . It is straightforward to show (see the Appendix) that the MLE of the fault magnitude  $\mu$  is given by

$$\hat{\mu} = \left( \sum_{i=1}^{k^*} \tilde{f}^2(i) \right)^{-1} \sum_{i=1}^{k^*} e(t - k^* + i) \tilde{f}(i), \tag{10}$$

which is very similar in form to (8). In Apley [26] it was shown that the estimate of the fault magnitude is unbiased, consistent, and efficient (i.e., achieves the Cramer-Rao lower bound). Regarding fault classification, suppose that both the step mean shift of (3) and a temporary spike (defined as  $f(t) = 0$  for  $t \neq \tau$  and  $f(t) = 1$  for  $t = \tau$ ) are to be simultaneously tested for. The fault signatures for each, which usually differ considerably, must be calculated using (6). The set of GLR statistics in (8) would be calculated for both faults and  $G(t)$  would be calculated as in (9), where the maximization is over both sets of GLR statistics. In general, the number of GLR statistics is the number of hypothesized faults multiplied by the window length. In complex production processes where different subsystems can experience different types of faults, the GLRT could prove to be a valuable diagnostic tool for classifying faults and tracking down root causes.

#### 4. Comparison of residual-based tests

This section compares the performance of three residual-based tests for detecting step mean shifts in the original process. The three tests are the GLRT of the previous section, a Shewhart individuals test applied to the residuals, and a two-sided CUSUM test applied to the residuals. We do not include a discussion of the EWMA test. We have found that the performance of the EWMA was always very similar to CUSUMs with appropriately chosen offset values, but never better than the optimally designed CUSUM (see also van der Wiel [16]). Consequently, it has not been included in the analysis.

##### 4.1. Three residual-based tests

The GLRT was discussed in detail in Section 3. The remaining two tests will be briefly outlined in this section.

*Shewhart individuals test.* The Shewhart test is the simplest to implement. At each time  $t$ ,  $|e(t)|$  is compared with a threshold  $H$ . If  $|e(t)|$  exceeds the threshold, a fault is signaled. Since  $e(t)$  is i.i.d. and Gaussian if no fault is present, a desired in-control ARL can be achieved by setting  $H$  equal to the  $1/(2 \times \text{ARL})$  quantile of the standard normal distribution.

*Two-sided CUSUM test.* The two-sided CUSUM test monitors for both positive and negative mean shifts. It involves tabulating two statistics – an upper CUSUM

statistic  $S_H(t)$  and a lower CUSUM statistic  $S_L(t)$ , defined by

$$S_H(t) = \max\{0, S_H(t-1) + e(t) - K\}, \text{ and}$$

$$S_L(t) = \max\{0, S_L(t-1) - e(t) - K\},$$

where  $K$  is an offset value. At each time  $S_H(t)$  and  $S_L(t)$  are compared to a threshold  $H$ , and if either falls above the threshold a fault is signaled.

Since this is a two degree-of-freedom design problem,  $H$  and  $K$  can be chosen to optimize some measure of the detection performance, while maintaining a desired in control ARL. Typical values of  $K$  range from  $0.25\sigma_a$  to  $1.5\sigma_a$ . For uncorrelated data experiencing step mean shifts, if one particular mean shift magnitude  $\mu$  is of interest, then a widely used rule-of-thumb is to select  $K = \mu/2$ . This choice of  $K$  very nearly minimized the ARL for a mean shift of magnitude  $\mu$ , under the constraint of a specified in control ARL [27]. However, when monitoring the residuals of an autocorrelated process, no such simple rule-of-thumb exists. One may consider setting  $K = \mu\tilde{f}_{ss}/2$ , but this may be far from the optimal value for a mean shift of magnitude  $\mu$ . By inspection of Fig. 1, in particular Models 1 through 4, one may surmise that the Shewhart test will perform better than the CUSUM with moderate values of  $K$ . Since the CUSUM becomes a Shewhart test in the limit, as  $K$  approaches the Shewhart control limit and  $H$  approaches 0, it is reasonable that values of  $K$  larger than  $\mu\tilde{f}_{ss}/2$  provide better detection performance. The following section demonstrates that this is, in fact, the case. Note that  $\tilde{f}_{ss}$  can be obtained directly from (7) with  $f_{ss} = 1$  for unit steps.

#### 4.2. Test comparison

The criteria for comparison is the probability of detecting step mean shifts of various magnitudes after a specified number of timesteps have passed – in most cases 20 timesteps. The notation  $P_{20}(\mu)$  will denote the probability of detecting a mean shift of magnitude  $\mu$  on or before the 20<sup>th</sup> timestep after its occurrence, and  $P_{10}(\mu), P_5(\mu)$ , etc. are similarly defined. The parameters of the tests were selected so that all have the same in-control ARL, denoted  $ARL(0)$ . Because of the forecast recovery phenomenon, the out-of-control ARLs tend to be disproportionately large when compared with the  $P_\bullet(\mu)$  values. Specifically, they are larger than what would result from analogous tests for mean shifts in uncorrelated processes with the same  $P_{10}(\mu)$  or  $P_{20}(\mu)$  values. Consequently, following the suggestion of Superville and Adams [15], we have chosen to use  $P_\bullet(\mu)$  instead of the out of control ARL as the means of comparison. See Superville and Adams [15] for a more elaborate discussion on forecast recovery.

For calculating the  $ARL(0)$  and  $P_\bullet(\mu)$  values a combination of theoretical derivation, Markov chain approximation and Monte Carlo approximation was used.

Theoretical derivation [17] was used for all Shewhart tests.  $ARL(0)$  for the CUSUM tests were estimated using the Markov chain method [24] applied to the one-sided version with the state-space discretized into 200 points. It was found that a finer discretization was not necessary for these examples. Since the two-sided CUSUM is symmetric and the  $ARL(0)$  values were relatively large, the two-sided ARL is very accurately approximated by one-half the one-sided ARL [28]. For discussions on alternatives to the Markov chain approach see Gan [29] and Woodall [25], for example.  $ARL(0)$  for the GLRT and the  $P_\bullet(\mu)$  values for the GLRT and CUSUM were all evaluated using Monte Carlo simulation with 20 000 trials. For  $ARL(0)$  for the GLRT, the simulations were truncated at 200 timesteps, with the last 100 being used to estimate the geometric decay constant of the tail of the run length distribution. The remaining tail of the distribution was then approximated using the estimated geometric decay constant.

The  $P_\bullet(\mu)$  comparison results for the six models of Table 1 and Fig. 1 are illustrated in Fig. 4. In each figure,  $P_\bullet(\mu)$  for the various tests is plotted as a function of  $\mu$ . In all cases  $P_{20}(\mu)$  values are shown, except for Model 5 for which  $P_{10}(\mu)$  values are shown.  $P_{10}(\mu)$  values are shown for Model 5 simply because for that model the detection probabilities for all of the tests were so high after 20 timesteps that it was difficult to draw meaningful conclusions. For all cases, the parameters of all tests were selected so that  $ARL(0)$  is 500. Similar analyses for  $ARL(0)$  set at 100 and 300 were also conducted, but the results were very similar and will not be shown.

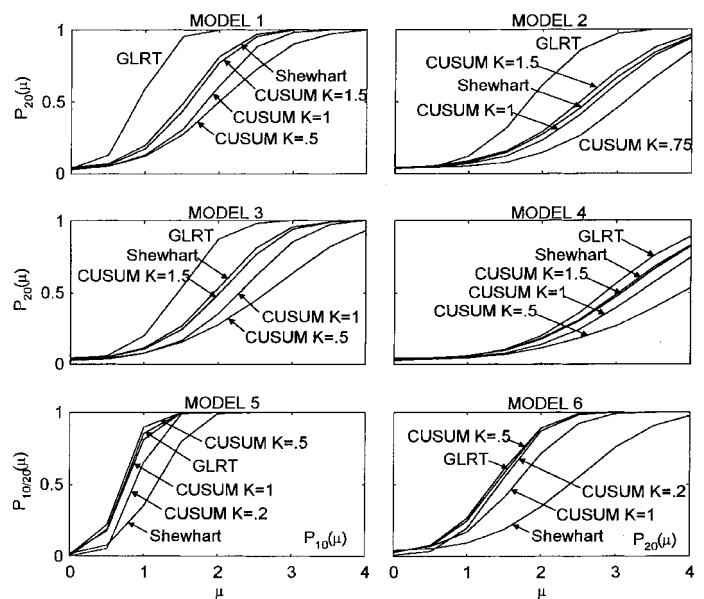


Fig. 4.  $P_{10}(\mu)$  values for various tests and the six models of Table 1.  $P_{20}(\mu)$  is shown for all models except Model 5, for which  $P_{10}(\mu)$  is shown. For all tests, the test parameters were chosen so that  $ARL(0) = 500$ .

For each of the models, CUSUM analyses were conducted with  $K$  taking values of  $\{0.2, 0.5, 0.75, 1.0, 1.25, 1.5, 2.0, 2.5\}$  and  $H$  taking corresponding values of  $\{9.96, 5.07, 3.54, 2.67, 2.11, 1.71, 1.11, 0.59\}$ . For each value of  $K$ , the corresponding value of  $H$  was chosen so that the resulting CUSUM has an  $ARL(0)$  of 500. For the sake of brevity CUSUMs with only selected values of  $K$  are illustrated in each figure. For each of the six models, the CUSUM with the  $K$  value that optimized its performance is always shown, unless the Shewhart outperformed all of the CUSUMs. In this event, the optimal CUSUM is the one with the largest  $K$  value and is nearly identical to the Shewhart. Strictly speaking, the Shewhart test can never be better than the optimal CUSUM, since the Shewhart is a special case of the CUSUM with  $K$  equal to the Shewhart control limit and  $H = 0$ . The values of  $K$  were chosen to cover the full range of values that would reasonably be used.  $K > 2.5$  is unnecessary, since CUSUM tests with values of  $K$  larger than that are nearly identical to the Shewhart test. It was also found that CUSUM tests with values of  $K < 0.2$  were never the optimal CUSUM for the examples used here.

For Models 1 through 3, the results are nearly identical. CUSUM tests with small values of  $K$  performed very poorly. Those with midrange values of  $K$  performed much better, but not quite as well as the Shewhart test or CUSUM tests with large  $K$ . Model 2 is an exception, where the optimal CUSUM ( $K = 1.5$ ) performed slightly better than the Shewhart test. The performance of the GLRT was significantly better than any of the other tests. This was expected, given the pronounced transient dynamics of the fault signatures for the first three models. Also, the results seem to hold uniformly for all values of  $\mu$  ranging from 0 to 4. For Model 4 the results are qualitatively the same, but the GLRT performance is only slightly better than the Shewhart test.

For Models 5 and 6, the conclusion is different. Here, the CUSUM with a midrange value of  $K$  (0.5) is the best test for both models. The Shewhart test performs poorly. The GLRT performs better than CUSUM tests with very small (0.2) and large (1.0)  $K$  values but slightly worse than the optimal CUSUM. From Fig. 1 it can be seen that there are very little transient dynamics in the fault signature for the GLRT to take advantage of.

It should be noted that the results for Models 5 and 6 were nearly identical qualitatively and quantitatively, except that all  $P_{\bullet}(\mu)$  values were scaled higher in Model 5 ( $f_{ss}$  is considerably larger). Consequently, the  $P_{10}(\mu)$  values for Model 5 can be used to draw conclusions about the  $P_{10}(\mu)$  values for Model 6 and *vice-versa* for the  $P_{20}(\mu)$  values. The main difference between  $P_{10}(\mu)$  and  $P_{20}(\mu)$  values is that the CUSUM tests with  $K = 0.2$  and  $K = 1$  change relative position and the performance of the Shewhart is more competitive after 10 timesteps than after 20. This is reasonable, since larger values of  $K$  result in smaller values of  $H$  and faster detection, providing the

mean shift is of sufficient size. The results for  $P_5(\mu)$  for Models 5 and 6 were along the same lines, with the CUSUM tests using  $K = 0.75$  (optimal) and  $K = 1$  performing slightly better than the GLRT, which performed slightly better than the CUSUM test with  $K = 0.5$ . The CUSUM with  $K = 0.2$  was worse than the Shewhart test. For all of the detection lags, the GLRT was quite close to the optimal CUSUM. In addition, no single CUSUM test outperformed the GLRT at all detection lags.

For Models 1 through 3 there was very little difference in the results when  $P_5(\mu)$  or  $P_{10}(\mu)$  were considered instead of  $P_{20}(\mu)$ . Moreover, the performance of the GLRT was still significantly better than any of the other tests. For Model 4 also there was little difference in the results, with one exception. When  $P_5(\mu)$  was used as the criteria, the Shewhart test had the best performance. Close behind were the CUSUM tests with  $K$  values larger than 1.5, with the GLRT close behind these.

The comparison results presented in this section can be reasonably generalized as follows. If the fault signature has significant transient dynamics (e.g., Models 1 through 3), in particular oscillation about zero, the GLRT outperforms the other tests by a wide margin. If the fault signature has very little transient dynamics and its steady-state magnitude is fairly large (e.g., Models 5 and 6), a CUSUM test with a properly chosen offset value is the best test, but only slightly better than the GLRT. For fault signatures with little transient dynamics and a small steady-state magnitude (e.g., Model 4), none of the tests perform very well. In these situations the GLRT may have a slight advantage over the Shewhart test, which may have a slight advantage over the CUSUM. The terms “significant” and “insignificant” transient dynamics are meant to be rough qualitative measures (rather than precise quantitative measures) of the transient dynamics, relative to the steady-state magnitude of the fault signature. To determine if a particular fault signature has significant transient dynamics, it can be compared to those of Fig. 1. The transient dynamics are significant for Models 1 through 3, marginal for Model 4, and insignificant for Models 5 and 6.

#### 4.3. The effects of modeling errors

As seen in the previous section, for many examples where CUSUM and Shewhart tests perform poorly, the GLRT performs significantly better. The explanation is simply that the GLRT makes more complete use of the model, both the ARIMA parameters and the shape of the fault. Thus, some comments on modeling errors are in order. While a thorough analysis of modeling errors is beyond the scope of this paper, the short analysis presented illustrates the main points. There are two types of modeling errors we consider: errors in the ARIMA parameters and errors in the assumed fault shape, discussed in order.



It is unreasonable to assume that the estimated AR-IMA model will exactly match the true model. Errors can result from a number of sources, most commonly inadequate data or nonstationarities, nonlinearities, or disturbances in the process. Tables 2 and 3 show the effects of ARIMA parameter errors on  $ARL(0)$  and  $P_{20}(\mu)$  for Models 2 and 6. The results for Models 1 and 3 were similar to Model 2, and those for Model 5 were similar to Model 6. It was assumed that the variance of the residuals (no longer  $\sigma_a^2$ ) was estimated correctly. For Model 2 (Table 2) it was assumed that the MA polynomial was incorrectly estimated as  $\hat{\Theta}(B)$ .  $\hat{\Theta}(B) = 1 - 0.28B + 0.64B^2$  represents a 10% increase in the magnitude of the roots of the MA polynomial, and  $\hat{\Theta}(B) = 1 - 0.62B + 0.81B^2$  represents a change in the phase angle of the MA roots from  $80^\circ$  to  $70^\circ$ . For Model 6 (Table 3), it was assumed the first order AR polynomial was estimated incorrectly as  $\hat{\Phi}(B)$  with approximately a 10% underestimation and a 10% overestimation of  $\Phi_1$ . For both models the effects depend strongly on the direction the parameters are misestimated.

From Tables 2 and 3 it is clear that errors in the AR-IMA parameters have strong effects on the performance of both the CUSUM and the GLRT. For situations like Model 6, where the fault signature has little dynamics and the GLRT performs similarly to the CUSUM, the effects of modeling errors are also similar. For situations like Model 2 where the GLRT performs dramatically better than the CUSUM, the GLRT is also more adversely affected by modeling errors. However, even with modeling errors the GLRT still appears to have considerably better performance than a Shewhart or CUSUM test for Model 2.

Table 3 shows that small errors in estimating  $\phi_1$  can either significantly increase or decrease  $ARL(0)$  for the CUSUM and GLRT, depending on whether  $\phi_1$  is over-

estimated or underestimated. For situations like Model 6, where there are little dynamics in the fault signature, the GLRT and CUSUM operate on similar principals: both test for general shifts away from zero in the average of a number of consecutive points. With  $\phi_1$  underestimated the residuals still have positive autocorrelation, and this event is more likely to occur even if there is no true mean shift. Thus,  $ARL(0)$ , is considerably decreased. With  $\phi_1$  overestimated the residuals have negative autocorrelation, and the exact opposite is true. In contrast, the Shewhart individuals test only tests individual points, and moderate autocorrelation in the residuals has little effect on  $ARL(0)$ .

The second type of modeling error to consider is mis-modeling of the shape of the fault  $f(t)$ . We consider the above situations where step mean shifts were assumed and are tested for, but the actual mean shift takes three timesteps to reach its final magnitude  $\mu$ . Specifically, in (2)  $f(1) = 1/3$ ,  $f(2) = 2/3$ , and  $f(j) = 1$ , for  $j > 2$ . It is assumed that there are no errors in the ARIMA parameters.  $P_{20}(\mu)$  results for the various tests are shown in Table 4 for Models 2, 4, and 6.  $ARL(0)$  is not shown since it is unaffected when the fault shape deviates from its assumed form. Results for Models 1 and 3 were conceptually very similar to those for Model 2. Results for Model 5 were similar to, although even less pronounced than, those of Model 6.

It is reasonable that the error in  $f(t)$  had very little effect for Models 5 and 6, given the shape of their fault signatures. Since there is very little dynamics to the fault signatures, all three tests rely predominantly on the size of  $\tilde{f}_{ss}$  to detect mean shifts, and  $\tilde{f}_{ss}$  was the same for both the assumed step fault and the actual fault. In contrast, errors in the fault shape have significant effects for Models 2 and 4. The tests which performed the best with no modeling errors are the most strongly affected. For

**Table 2.** Effects of errors in the ARIMA parameters for Model 2 on  $ARL(0)$  and  $P_{20}(\mu)$ . The true MA polynomial is  $\Theta(B) = 1 - 0.31B + 0.81B^2$ . The numbers shown in parentheses are the corresponding values when the true model is used

	$\hat{\Theta}(B) = 1 - 0.28B + 0.64B^2$		$\hat{\Theta}(B) = 1 - 0.62B + 0.81B^2$	
	$ARL(0)$	$P_{20}(2)$	$ARL(0)$	$P_{20}(2)$
GLRT	1123 (500)	0.238 (0.617)	130 (500)	0.556 (0.617)
CUSUM, $K = 0.75$	375 (500)	0.096 (0.144)	161 (500)	0.255 (0.144)
Shewhart	517 (500)	0.198 (0.273)	526 (500)	0.199 (0.273)

**Table 3.** Effects of errors in the ARIMA parameters for Model 6 on  $ARL(0)$  and  $P_{20}(\mu)$ . The true AR polynomial is  $\Phi(B) = 1 - 0.8B$ . The numbers shown in parentheses are the corresponding values when the true model is used

	$\hat{\Phi}(B) = 1 - 0.9B$		$\hat{\Phi}(B) = 1 - 0.7B$	
	$ARL(0)$	$P_{20}(1.5)$	$ARL(0)$	$P_{20}(1.5)$
GLRT	1087 (500)	0.158 (0.590)	162 (500)	0.831 (0.590)
CUSUM, $K = 0.5$	4223 (500)	0.122 (0.610)	133 (500)	0.854 (0.610)
Shewhart	492 (500)	0.118 (0.186)	479 (500)	0.286 (0.186)

**Table 4.** Effects of deviation from a step mean shift on  $P_{20}(\mu)$  values for various models and tests. The numbers in parentheses are the corresponding values with a step mean shift

	$P_{20}(2), Model 2$	$P_{20}(3), Model 4$	$P_{20}(1.5), Model 6$
GLRT	0.100 (0.617)	0.178 (0.566)	0.543 (0.590)
Shewhart	0.072 (0.273)	0.112 (0.494)	0.149 (0.186)
CUSUM, $K = 0.2$	0.010 (0.011)	0.156 (0.170)	0.517 (0.556)
CUSUM, $K = 0.5$	0.044 (0.063)	0.244 (0.267)	0.582 (0.610)
CUSUM, $K = 0.75$	0.058 (0.144)	0.235 (0.317)	0.476 (0.506)
CUSUM, $K = 1.0$	0.073 (0.234)	0.209 (0.392)	0.382 (0.411)
CUSUM, $K = 1.5$	0.081 (0.294)	0.157 (0.478)	0.227 (0.275)

Model 2, the GLRT is affected the most, followed by the Shewhart and CUSUM with large  $K$  values. CUSUMs with small  $K$  values, which performed dismally with no modeling errors, were much less affected. Even with the modeling errors considered, however, the GLRT performed better than any of the other tests. For Model 4 the GLRT, Shewhart, and CUSUM with  $K = 1.5$  are all affected similarly. CUSUMs with midrange  $K$  values are moderately affected, and the CUSUMs with small  $K$  are affected very little.

These results have an intuitive interpretation in terms of the shape of the fault signatures. Refer, for example, to the fault signature for Model 4 in Fig. 1. The GLRT, Shewhart, and CUSUMs with large  $K$  all rely heavily on the initial spike in the fault signature in order to detect the fault. When the mean shift is not a step, but instead takes three timesteps to build up, the initial spike in the fault signature is missing. For  $\mu = 1$ , the true fault signature ranges from 0.33 to 0.4 for the first three timesteps, after which it remains at 0.1. Thus, these tests are strongly affected. In contrast, CUSUMs with smaller  $K$  rely more on the steady-state magnitude of the fault signature than on initial dynamics to detect mean shifts. If the initial spike in the fault signature is missing, their performance does not suffer nearly as much as the GLRT or Shewhart. It should be kept in mind that, with no modeling errors, CUSUMs with small  $K$  performed the poorest of all the tests.

In the short robustness analysis presented in this section we have considered some simple estimation errors in order to illustrate the point that the tests can be sensitive to modeling errors. Parameter estimation errors and errors in the fault shape were considered separately. If such errors are considered together or if larger errors are considered, the results could be more extreme so that the Shewhart (for example) may be the best test. A more thorough robustness investigation would be very useful.

A final comment on the shape of the fault is in order. Even if the true shape of the fault is not a step, when the shape is known the GLRT can be modified accordingly. It is straightforward to develop the GLRT for mean shifts that, for example, build up linearly or exponentially, are sinusoidal, or are of any other form, providing they are repeatable. If the faults are repeatable and data

from prior occurrences of the fault is available, the shape of the fault can be estimated. The fault signature can then be calculated using (6), and (8) and (9) can be used directly.

## 5. Summary and conclusions

For detecting mean shifts in certain types of ARIMA processes, the GLRT presented in this paper can achieve dramatically improved performance over more conventional residual-based tests. The GLRT is quite simple to implement, requiring only that the fault signature (calculated off-line) be “correlated” with the residuals. In addition to any potential performance improvements, the GLRT possesses several other desirable features. It provides maximum likelihood estimates of the magnitude and time of occurrence of the fault and can be used to classify faults according to type, if more than one type is to be tested for. Thus, the GLRT could serve as a powerful diagnostic tool for detecting and isolating root causes in complex production processes that experience a number of different fault types.

Throughout this paper we have emphasized the importance of the fault signature. In addition to the direct use of the fault signature by the GLRT, it provides a simple yet effective means of determining which residual-based test will have the best performance, as the GLRT is not always the best test. By simple inspection of the fault signatures in Fig. 1, one could have guessed the results of the performance comparisons in Section 4: (i) for ARIMA models whose fault signatures have small steady-state magnitude, but pronounced transient dynamics, the GLRT outperforms the other tests by a wide margin, followed by the Shewhart test, with CUSUM tests performing poorly; (ii) if the fault signature has a relatively large steady-state magnitude and insignificant transient dynamics, a properly designed CUSUM test is as good, or slightly better, than the GLRT, and the Shewhart test performs poorly; and (iii) if the fault signature has neither a large steady-state magnitude, nor pronounced transient dynamics, none of the tests perform very well.

Given that the strength of the GLRT lies in its more complete use of the model, one might consider the pos-

sibility of modeling errors to be a deterrent to using the GLRT over, for example, a CUSUM test on the residuals. The brief analysis of modeling errors presented in Section 4 indicates that in cases where the CUSUM and GLRT perform comparably, they suffer comparably from modeling errors. The GLRT appears to be more adversely affected by modeling errors only in situations where, when the model is accurate, the GLRT drastically outperforms the CUSUM. For the examples studied where this was the case, after the modeling errors are taken into account, the GLRT still performed comparably to or slightly better than the CUSUM.

### Acknowledgements

This research was partially funded by NSF CAREER Grant: DMI 9624402. The authors are grateful to the editor and anonymous reviewers for their insightful comments, which have significantly improved this paper.

### References

- [1] Johnson, R.A. and Bagshaw, M. (1974) The effect of serial correlation on the performance of CUSUM tests. *Technometrics*, **16** (1), 103–112.
- [2] Vasilopoulos, A.V. and Stamboulis, A.P. (1978) Modification of control chart limits in the presence of data correlation. *Journal of Quality Technology*, **10** (1), 20–30.
- [3] Yashchin, E. (1993) Performance of CUSUM control schemes for serially correlated observations. *Technometrics*, **35** (1), 37–52.
- [4] Alwan, L.C. and Radson, D. (1992) Time-series investigation of subsample mean charts. *IIE Transactions*, **24**, 66–80.
- [5] Runger, G.C. and Willemain, T.R. (1996) Batch-means control charts for autocorrelated data. *IIE Transactions*, **28** (6), 483–487.
- [6] Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994) *Time Series Analysis: Forecasting and Control*, 3rd edn., Prentice Hall, Englewood Cliffs, N.J.
- [7] Alwan, L.C. and Roberts, H.V. (1988) Time-series modeling for statistical process control. *Journal of Business and Economic Statistics*, **6** (1), 87–95.
- [8] Berthouex, P.M., Hunter, W.G. and Pallesen, L. (1978) Monitoring sewage treatment plants: some quality control aspects. *Journal of Quality Technology*, **10**, 139–149.
- [9] Chow, M.C., Wu, S.M. and Ermer, D.S. (1979) A time series control chart for a nuclear reactor, in *Proceedings of the 1979 Annual Reliability and Maintainability Symposium*, IEEE Press, New York, NY, pp. 1–7.
- [10] English, J.R., Krishnamurthi, M. and Sastri, T. (1991) Quality monitoring of continuous flow processes. *Computers and Industrial Engineering*, **20** (2), 251–260.
- [11] Montgomery, D.C. and Mastrangelo, C.M. (1991) Some statistical process control methods for autocorrelated data. *Journal of Quality Technology*, **23** (3), 179–193.
- [12] Hu, S.J. and Roan, C. (1996) Change patterns in time series-based control charts. *Journal of Quality Technology*, **28** (3), 302–312.
- [13] Lin, W.S. and Adams, B.M. (1996) Combined control charts for forecast-based monitoring schemes. *Journal of Quality Technology*, **28** (3), 289–301.
- [14] Runger, G.C., Willemain, T.R. and Prabhu, S. (1995) Average run lengths for CUSUM control charts applied to residuals. *Communications in Statistics: Theory and Methods*, **24** (1), 273–282.
- [15] Superville, C.R. and Adams, B.M. (1994) An evaluation of forecast-based quality control schemes. *Communications in Statistics: Simulation and Computation*, **23** (3), 645–661.
- [16] van der Wiel, S.A. (1996) Monitoring processes that wander using integrated moving average models. *Technometrics*, **38** (2), 139–151.
- [17] Wardell, D.G., Moskowitz, H. and Plante, R.D. (1994) Run-length distributions of special-cause control charts for correlated processes. *Technometrics*, **36** (1), 3–17.
- [18] Willsky, A.S. (1976) A survey of design methods for failure detection in dynamic systems. *Automatica*, **12**, 601–611.
- [19] Apley, D.W. and Shi, J. (1994) A statistical process control method for autocorrelated data using a GLRT, in *Proceedings of the International Symposium on Manufacturing Science and Technology for the 21st Century*, Tsinghua Press, Beijing, China, pp. 165–170.
- [20] English, J.R. and Case, K.E. (1990) Control charts applied as filtering devices within a feedback control loop. *IIE Transactions*, **22** (3), 255–269.
- [21] MacGregor, J.F. and Harris, T.J. (1993) The exponentially weighted moving variance. *Journal of Quality Technology*, **25** (2), 106–118.
- [22] Oppenheim, A.V. and Schaffer, R.W. (1975) *Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ.
- [23] Basseville, M. (1988) Detecting changes in signals and systems – a survey. *Automatica*, **24** (3), 309–326.
- [24] Brooks, D. and Evans, D.A. (1972) An approach to the probability distribution of CUSUM run lengths. *Biometrika*, **59** (3), 539–549.
- [25] Woodall, W.H. (1983) The distribution of the run length of one-sided CUSUM procedures for continuous random variables. *Technometrics*, **25** (3), 295–301.
- [26] Apley, D.W. (1997) Supervisory adaptive control: monitoring, diagnostics and model uncertainty. Ph.D. dissertation, The University of Michigan, Ann Arbor, MI.
- [27] Montgomery, D.C. (1996) *Introduction to Statistical Quality Control*, 3rd edn., Wiley, New York.
- [28] van Dobben de Bruyn, C.S. (1968) *Cumulative Sum Tests: Theory and Practice*, Hafner, New York.
- [29] Gan, F.F. (1993) The run length distribution of a cumulative sum control chart. *Journal of Quality Technology*, **25** (3), 205–215.

### Appendix

This Appendix provides a brief derivation of the GLRT of (8) and (9). If the mean shift magnitude  $\mu$  were known *a priori*, a Likelihood Ratio Test (LRT) could be used. The LRT consists of, at each time  $t$ , choosing the hypothesis which maximizes the likelihood ratios

$$\Lambda_i \equiv \frac{p(\mathbf{e}|H_i, \mu)}{p(\mathbf{e}|H_0)} \quad (i = 1, 2, \dots, N). \quad (\text{A1})$$

Here,  $\mathbf{e} \equiv [e(t - N + 1)e(t - N + 2) \dots e(t)]^T$  is the vector of residuals over the current data window,  $H_i$  ( $i = 1, 2, \dots, N$ ) is the hypothesis that the fault has occurred at time  $t - i + 1$ ,  $H_0$  is the null hypothesis that no fault has occurred, and  $p(e|H_i, \mu)$  is notation for the conditional probability density of  $\mathbf{e}$ , given the  $i$ th hypothesis is true and the fault magnitude is  $\mu$ . Define  $\mathbf{f}_i \equiv [0 \dots 0 \tilde{f}(1) \tilde{f}(2) \dots \tilde{f}(i)]$ , where  $\tilde{f}(\bullet)$  is the fault signature for a fault occurring at timestep 1. It follows directly from (5) that, under  $H_i$ , the mean vector and covariance matrix of the jointly Gaussian random vector

$\underline{\mathbf{e}}$  are given by  $\mu\tilde{\mathbf{f}}_i$  and  $\sigma_a^2\mathbf{I}$ , where  $\mathbf{I}$  is the  $N \times N$  identity matrix. Substituting into (A1) the multivariate Gaussian probability density with these values for the mean and covariance matrix gives

$$\Lambda_i = \exp \left\{ \frac{1}{2\sigma_a^2} \left[ 2\mu\tilde{\mathbf{e}}^T\tilde{\mathbf{f}}_i - \mu^2\tilde{\mathbf{f}}_i^T\tilde{\mathbf{f}}_i \right] \right\}. \quad (\text{A2})$$

Since the fault magnitude  $\mu$  will not be known *a priori*, the GLRT approach is to substitute for  $\mu$  in (A2) its Maximum Likelihood Estimate (MLE) under  $H_i$ . The MLE of  $\mu$  under  $H_i$ , denoted  $\hat{\mu}_i$ , is given by

$$\begin{aligned} \hat{\mu}_i &= \arg \max_{\mu} \{ p(\underline{\mathbf{e}}|H_i, \mu) \} \\ &= \arg \min_{\mu} \{ [\underline{\mathbf{e}} - \mu\tilde{\mathbf{f}}_i]^T [\underline{\mathbf{e}} - \mu\tilde{\mathbf{f}}_i] \} = \frac{\underline{\mathbf{e}}^T\tilde{\mathbf{f}}_i}{\tilde{\mathbf{f}}_i^T\tilde{\mathbf{f}}_i}, \quad (\text{A3}) \end{aligned}$$

where the second equality follows from the definition of the multivariate Gaussian probability density and the fact that the mean vector and covariance matrix of  $\underline{\mathbf{e}}$  under  $H_i$  are  $\mu\tilde{\mathbf{f}}_i$  and  $\sigma_a^2\mathbf{I}$ .

Maximizing (A2) with (A3) substituted for  $\mu$  is equivalent to maximizing  $(\sigma_a^2\tilde{\mathbf{f}}_i^T\tilde{\mathbf{f}}_i)^{-1}(\underline{\mathbf{e}}^T\tilde{\mathbf{f}}_i)^2$ , which is the square of (8). Maximizing over  $i$  gives the GLRT of (9).

## Biographies

Daniel W. Apley received B.S. and M.S. degrees in Mechanical Engineering, an M.S. degree in Electrical Engineering, and a Ph.D. degree in Mechanical Engineering in 1990, 1992, 1995, and 1997, respectively, all from the University of Michigan. He was a visiting researcher at the

Institute for Machine Tools and Factory Management, Technical University of Berlin, in 1995. Since 1998, he has been with the Texas A&M University, where he is currently an Assistant Professor of Industrial Engineering. Dr. Apley's research area is quality assurance in manufacturing processes through in-process sensing and model-based statistical process monitoring, diagnosis, and adaptive feedback compensation. His recent research has been applied to chatter monitoring and suppression in machining, multivariate statistical process control and root cause diagnosis in autobody assembly, and active balancing of rotating machinery. He received an AT&T Bell Laboratories Ph.D. Scholarship from 1993 to 1997 and is a member of IIE, IEEE, ASME, and SME.

Jianjun (Jan) Shi is an Assistant Professor in the Department of Industrial and Operations Engineering at the University of Michigan. He received his B.S. and M.S. degrees in Electrical Engineering at the Beijing Institute of Technology in 1984 and 1987 respectively, and received his Ph.D. in Mechanical Engineering at the University of Michigan in 1992. Professor Shi's research interests are the fusion of advanced statistics and engineering knowledge to develop In-Process Quality Improvement (IPQI) methodologies achieving automatic process monitoring, diagnosis, compensation, and their implementation in various manufacturing processes. He has supervised six Ph.D. graduates and published more than 40 papers in this area. He received the NSF CAREER award in 1996, the Departmental Excellent Research Award at the University of Michigan in 1997, and the 1938E Award of the College of Engineering at the University of Michigan in 1998. His current research activities are sponsored by the General Motors Corporation, the Chrysler Corporation, Auto Body Consortium, National Institute of Standards and Technology – Advanced Technology Program, and the National Science Foundation. He serves as the director of the IPQI Research Lab in the IOE Department, the Associate Director of the S. M. Wu Manufacturing Research Center, and Program Technical Director at Auto Body Consortium. He is currently serving on the Editorial Board of *IIE Transactions on Quality and Reliability*. He is a member of ASME, ASQC, IIE, and SME.

*Contributed by the On-line Quality Engineering Department*