2002

Transfer Function Modeling of Processes with Dynamic Inputs

David West

Scott Dellana

See next page for additional authors

Citation/Publisher Attribution
David West, Scott Dellana, and Jeffrey Jarrett. 2002. "Transfer Function Modeling of Processes with Dynamic Inputs. *Journal of Quality Technology* 34, no. 3 (July 2002): 315-326. https://doi.org/10.1080/00224065.2002.11980161
Available at: https://doi.org/10.1080/00224065.2002.11980161

Authors
David West, Scott Dellana, and Jeffrey E. Jarrett

Transfer Function Modeling of Processes With Dynamic Inputs

DAVID WEST and SCOTT DELLANA

East Carolina University, Greenville, NC 27858-4353

JEFFREY JARRETT

The University of Rhode Island, Kingston, RI 02881

Time series structures, which are common occurrences with data in many industrial processes, complicate a quality practitioner's efforts to accurately position control chart limits. ARIMA modeling and a variety of control charting methods have been recommended for monitoring process data with a time series structure. Estimates of ARIMA model parameters may not be reliable, however, if assignable causes of variation are present in the process data used to fit the time series model. Control limits may also be misplaced if the process inputs are dynamic and exhibiting a time series structure. Our purpose in this paper is to explore the ability of a transfer function model to identify assignable causes of variation and to model dynamic relationships between process inputs and outputs. A transfer function model is developed to monitor biochemical oxygen demand output from a wastewater treatment process, a process with dynamic inputs. This model is used to identify periods of disturbance to the wastewater process and to capture the relationship between the variable nature of the input to the process and the resulting output. Simulation results are included in this study to measure the sensitivity of transfer function models and to highlight conditions where transfer function modeling is critical.

Introduction

IN an extensive survey, Alwan and Roberts (1995) found that more than 85% of industrial process control applications resulted in charts with possibly misplaced control limits. In many instances, the misplaced control limits result from autocorrelation of the process observations, which violates a basic assumption often associated with the Shewhart chart (Woodall (2000)). Autocorrelation of process observations has been reported in many industries, including cast steel (Alwan (1992)), blast furnace operations (Notohardjono and Ermer (1986)), wastewater treatment plants (Berthouex, Hunter, and Pallesen (1978)), chemical processes industries (Montgomery and Mastrangelo (1991)), semiconductor manufacturing (Kim and May (1994)), injection molding (Smith (1993)), and basic rolling operations (Xia, Rao, Shan, and Shu (1994)).

Several models have been proposed to monitor processes with autocorrelated observations. Alwan and Roberts (1988) suggest using an autoregressive integrated moving average (ARIMA) residuals chart, which they referred to as a special cause chart. For subsample control applications, Alwan and Radson (1992) describe a fixed limit control chart, where the original observations are plotted with control limit distances determined by the variance of the subsample mean series. Montgomery and Mastrangelo (1991) use an adaptive exponentially weighted moving average (EWMA) centerline approach, where the control limits are adaptive in nature and determined by a smoothed estimate of process variability. Lu and Reynolds (2001) investigate the steady state average run length of cumulative sum (CUSUM), EWMA, and Shewhart control charts for autocorrelated data

---

Dr. West is an Assistant Professor in the Department of Decision Sciences. He is a Member of ASQ. His email address is westd@mail.ecu.edu

Dr. Dellana is an Associate Professor in the Department of Decision Sciences. He is a Member of ASQ.

Dr. Jarrett is a Professor in the Department of Management Science and Statistics.

modeled as a first order autoregressive process plus an additional random error term.

A problem with all of these control models is that the estimate of the process variance is sensitive to outliers. If assignable causes are present in the data used to fit the model, the model may be incorrectly identified and the estimators of model parameters may be biased, resulting in loose or invalid control limits (Boyles (2000)). To justify the use of these methods, researchers have made the assumption that a period of "clean data" exists to estimate control limits. Therefore, methods are needed to assure that parameter estimates are free of contamination from assignable causes of variation. Intervention analysis, with an iterative identification of outliers, has been proposed for this purpose. The reader interested in more detail should see Alwan (2000, pp. 301-307), Atienza, Tang, and Ang (1998), and Box, Jenkins, and Reinsel (1994, pp. 473-474). Atienza, Tang, and Ang (1998) recommend the use of a control procedure based on an intervention test statistic, λ, and show that their procedure is more sensitive than ARIMA residual charts for process applications with high levels of positive autocorrelation. They limit their investigation of intervention analysis, however, to the detection of a single level disturbance in a process with high levels of first order autocorrelation. Wright, Booth, and Hu (2001) propose a joint estimation method capable of detecting outliers in an autocorrelated process where the data available is limited to as few as 9 to 25 process observations. Since intervention analysis is crucial to model identification and estimation, we investigate varying levels of autocorrelation, autoregressive and moving average processes, different types of disturbances, and multiple process disturbances.

The ARIMA and intervention models are appropriate for autocorrelated processes whose input streams are closely controlled. However, there are quality applications, which we refer to as "dynamic input processes," where this is not a valid assumption. The treatment of wastewater is one example of a dynamic process that must accommodate highly fluctuating input conditions. In the health care sector, the modeling of emergency room service must also deal with highly variable inputs. The dynamic nature of the input creates an additional source of variability in the system, namely the time series structure of the process input. For these applications, modeling the dynamic relationship between process inputs and outputs can be used to obtain improved process monitoring and control as discussed by Alwan (2000, pp. 675-679).

We propose a more general transfer function: an ARIMA model that accounts for both outliers in process output and dynamic effects from process input. In the following section, we briefly describe the relevant theory of time series analysis used in this paper. We then analyze the transfer function model terms to identify disturbances in a wastewater treatment process. We follow in later sections with supporting empirical evidence on the sensitivity of these methods. The paper concludes with a discussion of the implications for quality practitioners who may be monitoring processes which produce data with time series structures and which have dynamic inputs. In this paper, we refer to autocorrelated processes as either autoregressive (AR) or as moving average (MA) (as defined by Box, Jenkins, and Reinsel (1994)).

Transfer Function Modeling of Process Data

If a process quality characteristic, z_t, has a time series structure, an ARIMA model of the following general form can represent the undisturbed or natural process variation:

$$\phi(B)\alpha(B)z_t = \theta(B)a_t. \tag{1}$$

In Equation (1), B represents the back-shift operator, where $B(z_t) = z_{t-1}$. The value of $\phi(B)$ represents the polynomial expression $(1 - \phi_1 B - \ldots - \phi_p B^p)$, which models the autoregressive (AR) structure of the time series. The value of $\theta(B)$ represents the polynomial $(1 - \theta_1 B - \ldots - \theta_q B^q)$, which models the moving average (MA) structure of the time series. The value of $\alpha(B)$ represents the expression $(1-B)_1^d(1-B^s)_2^d$, where $d = d_1 + sd_2$. This quantity is a polynomial in B that expresses the degree of differencing required to achieve a stationary series and accounts for any seasonal pattern in the time series. Finally, a_t is a white noise series with distribution $N(0, \sigma_a^2)$. This model is described by Chen and Liu (1993b). If the series z_t is contaminated by periods of external disturbances to the process, the ARIMA model may be incorrectly specified, the variability of the residuals overestimated, and the resulting control limits incorrectly placed.

The following transfer function model of Box and Tiao (1973) describes the observed quality characteristic, y_t, as a function of three sources of variability:

$$y_t = \nu(B)x_{t-b} + \frac{\omega(B)}{\delta(B)}I_t + \frac{\theta(B)}{\phi(B)}a_t. \tag{2}$$

The first term, $\nu(B)x_{t-b}$, is the dynamic input term and represents an impulse function, $\nu(B)$, applied to the input x_{t-b} with a lag of b time periods. If a dynamic relationship between the input and output time series exists, lagged values of process inputs can be modeled, resulting in considerable reduction of unexplained variance. The second term, $(\omega(B)/\delta(B))I_t$, is the intervention term and identifies periods of time when assignable causes are present in the process. Here, I_t is an indicator variable with a value of zero when the process is undisturbed and a value of one when a disturbance is present in the process. See, for example, Box, Jenkins, and Reinsel (1994, p. 392) for the development of the transfer function term, and Box, Jenkins, and Reinsel (1994, p. 462) for details of the intervention term. The rational coefficient term of I_t is a ratio of polynomials that defines the nature of the disturbance as detailed in Box, Jenkins, and Reinsel (1994, p. 464). The third term, $(\theta(B)/\phi(B))a_t$, is the basic ARIMA model of the undisturbed process from Equation (1). We refer to Equation (2) as the "transfer function" model throughout this paper.

Different types of disturbances can be modeled by the proper design of the intervention term. The two most common disturbances for quality applications are a point disturbance, with an impact observed for only a single time period, and a step disturbance, with an impact persisting undiminished through several subsequent observations. The point disturbance is modeled as an additive outlier (AO). An AO impacts the observed process at one observation. The AO is modeled in the form

$$\frac{\omega(B)}{\delta(B)} = \omega_0, \tag{3}$$

where ω_0 is a constant. A step disturbance to the process is modeled as a level-shift outlier (a form of innovational outlier or IO) in the form,

$$\frac{\omega(B)}{\delta(B)} = \frac{\omega_0}{1-B}. \tag{4}$$

Chang, Tiao, and Chen (1988) and Chen and Liu (1993b) discuss both types of disturbance.

Chang, Tiao, and Chen (1988) extended the concepts of Box and Tiao (1975) to an iterative method for detecting the location and nature of outliers at unknown points in the time series. They define a procedure for detecting innovational outliers and additive outliers, and for jointly estimating time series parameters. Their work also demonstrates that the presence of outliers may cause serious bias in the estimation of ARIMA model parameters. The iterative method of Chang, Tiao, and Chen (1988) is effective for large, isolated outliers, but a masking effect may occur when multiple outliers cluster together. To overcome these problems, Chen and Liu (1993a) proposed an iterative outlier detection and adjustment methodology designed to handle multiple outliers of various types in a time series.

Box, Jenkins, and Reinsel (1994, p. 473) report the significance of the intervention term in the estimation of σ_a for two chemical process applications. For a chemical process temperature reading, three disturbances were identified by the intervention terms, resulting in a 26% reduction in $\widehat{\sigma}_a$. For chemical process viscosity readings, a single disturbance was identified by the intervention term and modeled with a resulting reduction in $\widehat{\sigma}_a$ of 6 percent.

Application of Transfer Function Modeling to Wastewater Treatment

Wastewater contains a variety of substances that must be removed, such as human wastes, food scraps, oils, soaps, chemicals, microorganisms, phosphorous compounds, nitrogen compounds, suspended solids, and organic matter. A critical step in the typical wastewater treatment process, shown in Figure 1, is the use of microorganisms to decompose the organic matter. The amount of oxygen consumed by the microorganisms, called the biochemical oxygen demand (BOD), is used as a process control measure. The challenges of controlling the BOD levels of a wastewater treatment plant result from the complexity of the process and the variances in the composition and flow rate of the input stream as discussed by Wen and Vassiliadis (1998). The value of transfer function modeling is illustrated in the following example, where we analyze 527 daily measurements of BOD from an urban wastewater treatment plant as reported by Poch, Bejar, and Cortes (1993). The dynamic nature of the input and output BOD is evident from the first 100 daily measurements plotted in Figures 2 and 3.

In the following subsections, we sequentially analyze the three sources of variability in the transfer function model of Equation (2). We first evaluate an ARIMA residual chart to serve as a benchmark. The second subsection details an analysis that includes intervention terms in the model, and the final subsection includes the dynamic input term. In all cases, model parameters are estimated from the

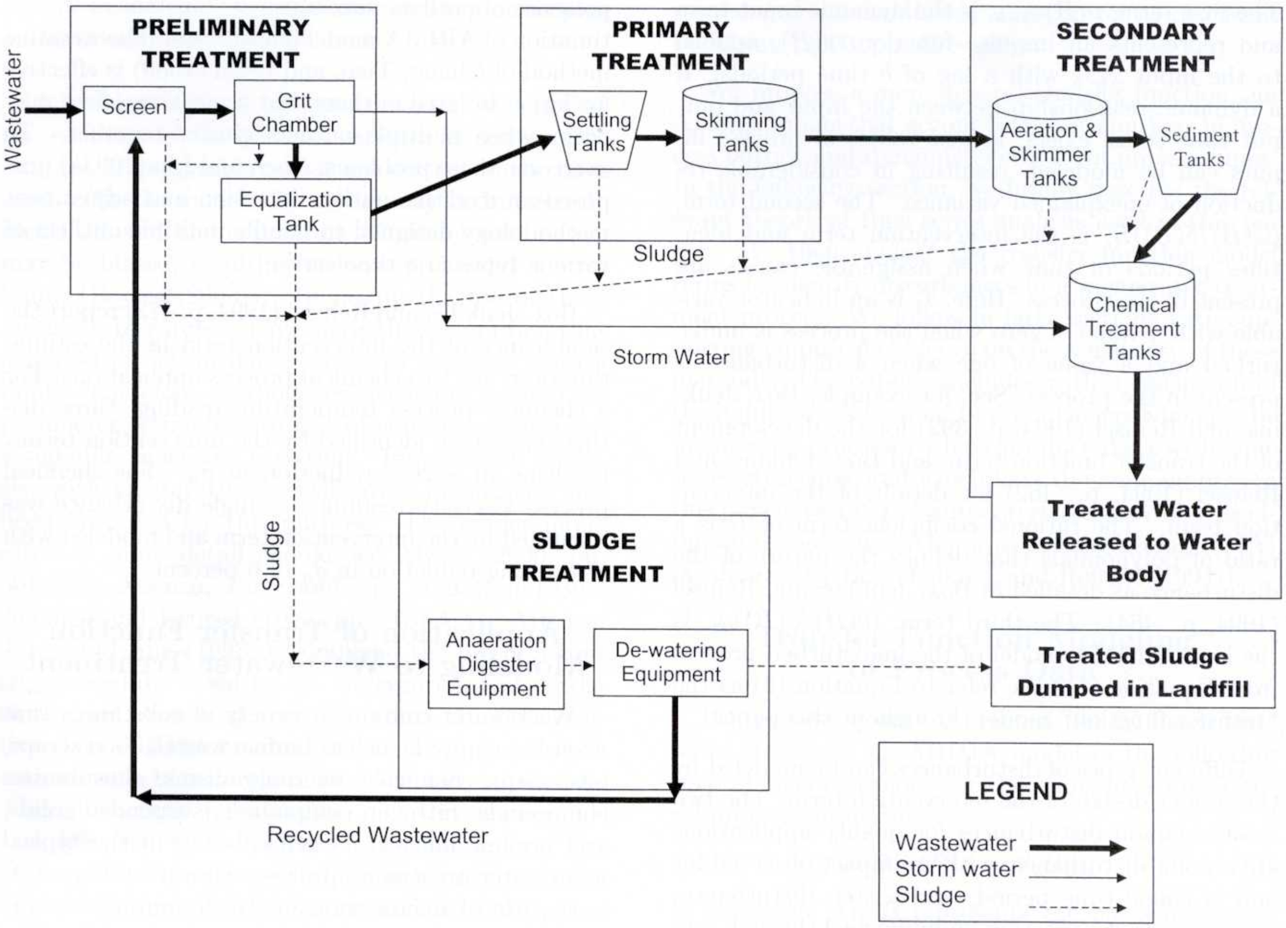

FIGURE 1. Typical Wastewater Treatment Plant.

first 100 days of operations and the remaining 427 days serve as an independent holdout sample. The ARIMA analysis and estimation was performed using SAS software; the intervention test statistic, λ, was calculated with an Excel spreadsheet using a method similar to the approach of Atienza, Tang, and Ang (1998).

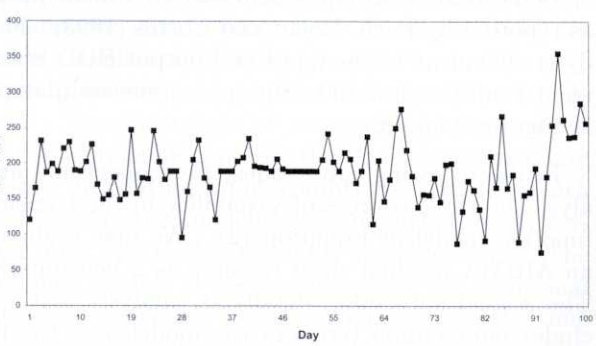

FIGURE 2. Wastewater Treatment Input BOD Time Series.

ARIMA Model Residual Analysis of the Wastewater Process

The construction of ARIMA residual control charts begins by analyzing the autocorrelation and partial autocorrelation functions to identify an appropriate ARIMA model for the BOD output. Box,

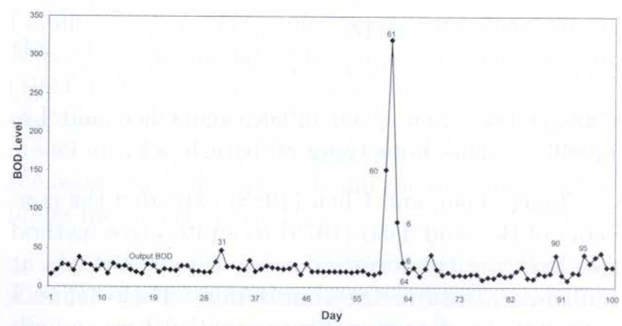

FIGURE 3. Wastewater Treatment Output BOD Time Series.

TABLE 1. Wastewater Control Models

Model	$\widehat{\sigma}_a$	Control Limits	Outliers Detected 1-100	Outliers Detected 101-527
ARIMA Residual	28.1	±84.3	60, 61	465
Intervention Term	6.24	±18.7	31, 60-64, 90, 95	465-467
Intervention and Dynamic Input Term	5.84	±17.5	31, 60-64, 90, 95	114, 465-467

Jenkins, and Reinsel (1994, pp. 183-223) and Montgomery and Johnson (1976, pp. 193-200) discuss such an analysis. The autocorrelation function for the first 100 observations cuts off abruptly at a lag of one, while the partial autocorrelation function follows a sinusoidal decay pattern. This information suggests that the process is stationary, and that an MA(1) model is appropriate for the BOD output of the wastewater treatment plant. Estimation of MA(1) model parameters yields the following result:

$$\widehat{y}_{t+1} = 27.47 + 0.582e_t. \tag{5}$$

If the fitted model is an adequate representation of the sample data, then the model residuals shown in Figure 4 should approximate a white noise process, an important property for the success of the residuals control chart. The obvious disturbance beginning in day 60 can affect model identification and parameter estimates. The magnitude of this impact will be clear in the next subsection, where we identify and model these events. For the moment, we will ignore the disturbance and estimate σ_a in Equation (5). The results, shown in Table 1, are that $\widehat{\sigma}_a = 28.1$ and that the 3-sigma limits are ± 84.3. The residuals control chart identifies days 60 and 61 of Figure 4 as potential process disturbances. Applying these control limits to holdout observations for days 101-527 identifies a single disturbance at day 465.

Addition of Intervention Analysis to the Wastewater Process Model

The use of the iterative method of Chang, Tiao, and Chen (1988) can improve on the ARIMA residuals results by identifying and explicitly modeling wastewater process disturbances. This method is based on the likelihood ratio criteria, λ, defined by Fox (1972), and uses the following algorithm:

1. Treat the observed time series, z_t, as if no outliers exist. Identify an ARIMA model and estimate model parameters (i.e. the polynomials defined in Equation (1)).
2. From the model estimated in step 1, compute the residuals, e_t, and $\widehat{\sigma}_a$.
3. Estimate λ_{1T} and λ_{2T} to determine the presence of AO and IO outliers, respectively, for $t = 1$ to n observations. The first statistic, λ_{1T}, is used to test the hypothesis that H_0: $\omega_0 = 0$, H_1: $\omega_0 \neq 0$ in Equation (3). The second statistic, λ_{2T}, is used to test H_0: $\omega_0 = 0$, H_2: $\omega_0 \neq 0$ in Equation (4). These estimators are defined as

$$\lambda_{1T} = \frac{e_T}{\widehat{\sigma}_a} \tag{6}$$

and

$$\lambda_{2T} = \frac{\bar{\omega}_A}{\rho\widehat{\sigma}_a}, \tag{7}$$

where

$$\rho^2 = (1 + \pi_1^2 + \pi_2^2 + \ldots + \pi_{n-T}^2)^{-1}, \tag{8}$$

and

$$\bar{\omega}_A = \rho^2 \pi(F)\pi(B)z_t. \tag{9}$$

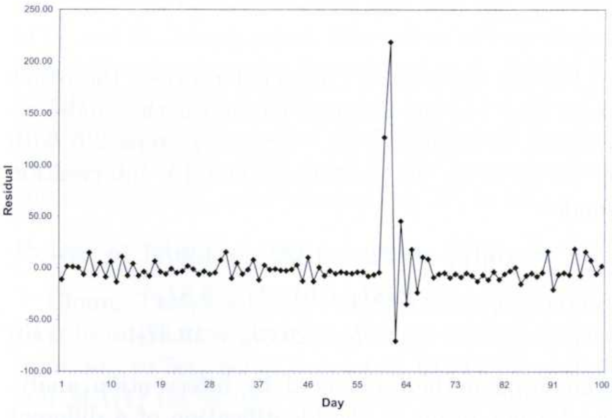

FIGURE 4. Wastewater Output BOD MA(1) Model Residuals.

In Equations (8) and (9), the π_i terms are the coefficients of the following rational polynomial with terms defined in Equation (1):

$$\pi(B) = \frac{\phi(B)\alpha(B)}{\theta_B}.$$

The symbol F represents the forward shift operator.

4. Find the maximum magnitude of the series of statistics λ_{1T} and λ_{2T}.

$$\eta_{IO} = \max|\lambda_{1,t}| \qquad \eta_{AO} = \max|\lambda_{2,t}|$$
$$t = 1, \ldots, n.$$

If either or both of these values exceeds a threshold, C, which typically varies from 2.5 to 4, it is determined that an outlier exists. If the maximum value is η_{IO}, then the outlier is an IO. Conversely, if the maximum is η_{AO}, then the outlier is an AO. The residuals are adjusted by a method that is dependent on the nature of the outlier. For an IO outlier, the model residual for that single period is set to zero. For an AO outlier the model residuals are adjusted using

$$e_{t,new} = e_t - \bar{\omega}_A \pi(B) I_T. \tag{12}$$

After adjusting the model residuals, a new value of $\widehat{\sigma}_a$ is computed.

5. Continue to identify additional outliers by repeating steps 3 and 4 with modified residuals and re-estimating σ_a until no further outlier candidates exist.

6. Treat the disturbance times for all outliers identified above as known and simultaneously estimate the ARIMA model parameters and interventions. Repeat steps 1 to 6 with this model and stop when no additional outliers are identified.

For the wastewater treatment process, the result from step 1 of the iterative method is the model estimated in Equation (5). Applying steps 2-6 with $C = 3$ yields the following ARIMA intervention model:

$$\begin{aligned} \widehat{y}_{t+1} &= 22.2 + 0.337e_t + 130I_1 \\ &\quad + 298I_2 + 61.8I_3 + 9.8I_4 \\ &\quad - 44.5I_5 + 20.3I_6 + 19.3I_7. \end{aligned} \tag{10}$$

The seven outliers detected by intervention analysis did not result in the identification of a different form of model; in this case the model remains an MA(1) model. However, the value of $\widehat{\theta}_1$ is reduced from 0.582 (estimated for the basic ARIMA residuals chart) to 0.337. Most importantly, the estimate $\widehat{\sigma}_a$, which is the basis for placing control limits, is reduced. The estimate $\widehat{\sigma}_a = 28.1$ of the basic ARIMA residuals is reduced significantly to $\widehat{\sigma}_a = 6.24$ with the inclusion of intervention terms (see Table 1). This results in tighter placement of control limits for a residuals chart and more sensitivity to identify additional potential disturbances to the wastewater process. Potential disturbances are now identified for days 31, 60–64, 90, and 95 and for days 465–467 of the holdout sample.

Addition of Dynamic Input Analysis to the Wastewater Process Model

We next include a term for the dynamic regression of the input BOD in the control model. Autocorrelation and partial autocorrelation functions calculated for the input BOD suggest that this time series is also an MA(1) process. We fit a transfer function model of the form of Equation (2) yielding the following result:

$$\begin{aligned} \widehat{y}_t &= 8.1 + .037x_t + .039x_{t-1} \\ &\quad + .228e_{t-1} + I_t, \end{aligned} \tag{11}$$

where x_t is the BOD input to the process at time t and I_t represents the same seven intervention terms defined in Equation (10). By modeling the variability of the BOD input to the wastewater process, the estimate $\widehat{\sigma}_a$ is reduced to 5.84 (see Table 1). The placement of the control limits is now at ± 17.5, the tightest control limits of all three models investigated. Including the dynamic input term in the model provides increased sensitivity, identifying a potential disturbance at day 114, which use of the model in Equation (10) did not detect.

Sensitivity of Intervention Analysis

The quality practitioner may be concerned with the power of intervention analysis to detect and model process disturbances and in understanding those process conditions where it is most important to use transfer models. We explore these issues next using simulated data.

Impact of Process Disturbances on Variance Estimation

To understand the impact of outliers on the estimation of the noise variance in ARIMA models, we estimate σ_a for simulated time series ranging in length from 50 to 200 observations for θ_1 and ϕ_1 values of 0.3, 0.5, and 0.7. The impact of a single point

TABLE 2. Impact of Single Disturbance on $\widehat{\sigma}_a$ for AR(1) Time Series Model

Distance Size w	Series Length n	AR Level ϕ_1	ARIMA $\widehat{\sigma}_a$	Control Limits $\pm 3\widehat{\sigma}_a$
1	50	0.3	1.01	3.03
2	50	0.3	1.04	3.12
3	50	0.3	1.10	3.30
4	50	0.3	1.14	3.42
1	50	0.5	1.01	3.03
2	50	0.5	1.04	3.12
3	50	0.5	1.09	3.27
4	50	0.5	1.15	3.45
1	50	0.7	1.01	3.03
2	50	0.7	1.05	3.15
3	50	0.7	1.10	3.30
4	50	0.7	1.15	3.45
1	125	0.5	1.00	3.00
2	125	0.5	1.01	3.03
3	125	0.5	1.03	3.09
4	125	0.5	1.06	3.18
1	200	0.3	1.00	3.00
2	200	0.3	1.01	3.03
3	200	0.3	1.02	3.06
4	200	0.3	1.04	3.12
1	200	0.5	1.00	3.00
2	200	0.5	1.01	3.03
3	200	0.5	1.02	3.06
4	200	0.5	1.04	3.12
1	200	0.7	1.00	3.00
2	200	0.7	1.01	3.03
3	200	0.7	1.02	3.06
4	200	0.7	1.04	3.12

disturbance of sizes 1, 2, 3, and 4 in units of σ_a on estimates of σ_a are measured. The results reported in Table 2 for AR(1) models and Table 3 for MA(1) models are averages of 1,000 repetitions. Since the simulated data a_t is constructed with $\sigma_a = 1$, the 3-sigma control limits should be at ± 3. To the extent that the model estimate of σ_a exceeds 1, the control limits are misplaced due to the failure to identify and model the process disturbance. The results suggest that for time series of length greater than 200 observations, a single disturbance of $4\sigma_a$ or less will have a modest impact of 4% or less on the placement of control limits. For shorter time series of approximately 50 observations, disturbances of $2\sigma_a$ or greater will significantly bias the estimators of control limits. These limits will be overestimated from 4% to 16%. These conclusions hold for both AR(1) and MA(1) time series structures and for all levels of θ_1 and ϕ_1 investigated. The effect of two disturbances in the time series is summarized in Table 4 for an AR(1) time series with $\phi_1 = 0.5$. In this case, the shorter time series control chart limits will be overestimated by 9% to 30%, and, for the longer series of 200 observations, overestimated by 2% to 9%.

Power of Iterative Intervention Analysis

Chang, Tiao, and Chen (1988) report the power of their iterative method to detect large outliers varying from $3\sigma_a$ to $5\sigma_a$ using simulated AR(1) ($\phi_1 = 0.6$) and MA(1) ($\theta_1 = 0.6$) time series consisting of 150 observations. They found that the iterative method achieves a power of 0.975 for identification of an IO,

TABLE 3. Impact of Single Disturbance on $\widehat{\sigma}_a$ for MA(1) Time Series Model

Disturbance Size w	Series Length n	AR Level θ_1	ARIMA $\widehat{\sigma}_a$	Control Limits $\pm 3\widehat{\sigma}_a$
1	50	0.3	1.00	3.00
2	50	0.3	1.04	3.12
3	50	0.3	1.09	3.27
4	50	0.3	1.15	3.45
1	50	0.5	1.01	3.03
2	50	0.5	1.04	3.12
3	50	0.5	1.09	3.27
4	50	0.5	1.14	3.42
1	50	0.7	1.01	3.03
2	50	0.7	1.05	3.15
3	50	0.7	1.09	3.27
4	50	0.7	1.16	3.48
1	125	0.5	1.01	3.03
2	125	0.5	1.01	3.03
3	125	0.5	1.03	3.09
4	125	0.5	1.06	3.18
1	200	0.3	1.00	3.00
2	200	0.3	1.01	3.03
3	200	0.3	1.02	3.06
4	200	0.3	1.04	3.12
1	200	0.5	1.00	3.00
2	200	0.5	1.01	3.03
3	200	0.5	1.02	3.06
4	200	0.5	1.04	3.12
1	200	0.7	1.00	3.00
2	200	0.7	1.01	3.03
3	200	0.7	1.02	3.06
4	200	0.7	1.04	3.12

and a power of 0.99 for identification of an AO in both AR(1) and MA(1) time series with $5\sigma_a$ disturbances. For the smaller $3\sigma_a$ disturbance, the power is reduced to 0.55 for an IO and 0.69 for an AO for both time series structures.

We extend the work of Chang, Tiao, and Chen (1988) by investigating the power of the iterative detection method for outliers of $1\sigma_a$, $2\sigma_a$, and $3\sigma_a$ in both AR(1) and MA(1) time series to characterize more typical process-related outliers. The experi-

TABLE 4. Impact of Two Disturbances on $\widehat{\sigma}_a$ for AR(1) Time Series Model

Disturbance Size w	Series Length n	AR Level ϕ_1	ARIMA $\widehat{\sigma}_a$	Control Limits $\pm 3\widehat{\sigma}_a$
2	50	0.5	1.09	3.27
4	50	0.5	1.30	3.90
2	125	0.5	1.03	3.03
4	125	0.5	1.14	3.42
2	200	0.5	1.02	3.06
4	200	0.5	1.09	3.27

TABLE 5. Factorial Design

Factor	Factor levels
AR (ϕ) or MA (θ) Level	0.3, 0.5, 0.7
Type of Assignable Cause	Step, Point
Size of Assignable Cause Disturbance	0.0, 1.0, 2.0, 3.0 (multiples of σ_a)

mental comparisons are based on a full factorial design with three design factors. The design factors, summarized in Table 5, are the level of the time series structure in the manufacturing process (represented by either ϕ_1 for the autoregressive model or θ_1 for the moving average model), the nature of the assignable cause of variation, and the magnitude of the disturbance. The parameters of the time series structure include a low level of 0.3, a moderate level of 0.5, and a high level of 0.7. Four different levels of disturbances are studied: $0.0\sigma_a$, $1.0\sigma_a$, $2.0\sigma_a$, and $3.0\sigma_a$. We investigate both point disturbances and step disturbances. The disturbance is generated at random locations between observation 2 and observation 150, which is the end of the simulated data. Reported results are averages of 1000 repetitions for each condition. The two aspects of the power of intervention models in process control applications are measured in this study by the rate of true positives and the rate of false positives (or false alarms).

TABLE 6. Iterative Intervention Analysis Results AR(1) Models (1000 Runs Per Experiment, $C = 3$)

AR Level (ϕ_1)	Type of Assignable Cause	Size of Special Cause Disturbances (multiples of σ_a)	Proportion of Disturbances Detected
0.30	None	0	.003
	Step	1	.007
		2	.038
		3	.201
0.50	None	0	.006
	Step	1	.011
		2	.057
		3	.265
0.70	None	0	.004
	Step	1	.008
		2	.074
		3	.309
0.30	None	0	.006
	Point	1	.018
		2	.154
		3	.505
0.50	None	0	.001
	Point	1	.031
		2	.171
		3	.618
0.70	None	0	.002
	Point	1	.028
		2	.228
		3	.702

TABLE 7. Iterative Intervention Analysis Results MA(1) Models (1000 Runs Per Experiment, $C = 3$)

AR Level (θ_1)	Type of Assignable Cause	Size of Special Cause Disturbances (multiples of σ_a)	Proportion of Disturbances Detected
0.30	None	0	.002
	Step	1	.010
		2	.016
		3	.055
0.50	None	0	.007
	Step	1	.005
		2	.042
		3	.089
0.70	None	0	.001
	Step	1	.014
		2	.058
		3	.175
0.30	None	0	.007
	Point	1	.021
		2	.147
		3	.455
0.50	None	0	.004
	Point	1	.012
		2	.139
		3	.452
0.70	None	0	.002
	Point	1	.018
		2	.124
		3	.400

The experimental results of the intervention analysis are reported in Table 6 for the AR(1) models and in Table 7 for the MA(1) models. These tables summarize the proportion of times that a disturbance was detected by the intervention analysis for each experimental condition. Our operational definition for detecting a disturbance is based on using a threshold value of $C = 3$. We also report operating characteristic curves for other values of the threshold C in Figure 5.

The results of the AR(1) time series of Table 6 indicate that a point disturbance is easier to identify than a step disturbance. This is likely caused by the fact that the persistent nature of the step disturbance results in significantly overestimating σ_a and thereby understates the values of λ calculated in Equations (6) and (7). The results also suggest that it is easier to detect a disturbance at higher values of autocorrelation. The $3\sigma_a$ point disturbance is detected 50.5% of the time at $\phi_1 = 0.3$, 61.8% at $\phi_1 = 0.5$, and 70.2% at $\phi_1 = 0.7$. The corresponding results for the step disturbance are 20.1% for $\phi_1 = 0.3$, 26.5% for $\phi_1 = 0.5$, and 30.9% for $\phi_1 = 0.7$. Detection of disturbances smaller than $3\sigma_a$ is unreliable. From 1% to 23% of disturbances of size $1\sigma_a$ and $2\sigma_a$ are detected.

From the results of Table 7, it is evident that outlier detection is more difficult with the MA(1) than

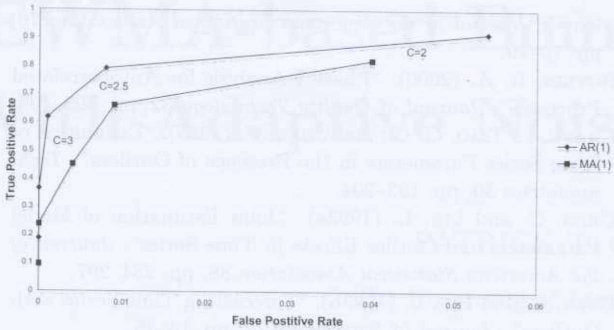

FIGURE 5. Operating Characteristics of Intervention Detection.

with AR(1) time series. For the MA(1) series, the $3\sigma_a$ point disturbance is detected 45.5% of the time at $\theta_1 = 0.3$, 45.2% at $\theta_1 = 0.5$, and 40% at $\theta_1 = 0.7$. Again, the smaller disturbances are not reliably identified, with detection proportions ranging from 0.5% to 15%.

The false positive rate can be inferred from Tables 6 and 7 by the proportion of disturbances detected for the experimental conditions with a zero disturbance size. Since there is no disturbance for this condition, the distinction between the step and point disturbance is irrelevant. By averaging across all three parameter levels, we estimate that the intervention model detected a "false disturbance" about 0.37% of the time. A similar analysis for the MA(1) results yields a false positive rate of 0.38%.

The selection of an appropriate threshold, C, to use in identifying potential outliers is dependent on the economics of the process. To guide this decision we plot operating characteristic curves in Figure 5 that define the tradeoff between true positive and false positive rates for the detection of a single disturbance in AR(1) and MA(1) time series with threshold values of $C = [2, 2.5, 3, 3.5, 4]$. Higher values of C correspond to the lower left of the figure and lower values to the upper right. For both AR(1) and MA(1) time series, a significant increase in sensitivity occurs by lowering the threshold from $C = 4$ to $C = 2.5$. Below $C = 2.5$, the gains in sensitivity are at the expense of a substantially higher number of false positives. At $C = 2.5$, approximately 79% of point disturbances are identified in AR(1) time series and 66% in MA(1) series. The corresponding false positive rate is slightly less than 1%.

Implications for Quality Practice

Transfer function modeling is shown to be useful in monitoring process quality because of its ability to identify process disturbances resulting from assignable sources of variation prior to the estimation of model parameters, and its ability to explicitly model relationships between dynamic process inputs and output quality levels. This allows the quality practitioner to more accurately estimate the process variability and minimize the problem of misplaced residual chart control limits. The use of a transfer function model is most beneficial for relatively short time series with 50 to 200 observations. For these short time series, control limits can be overestimated by as much as 15% from a single reasonably small point disturbance and overestimated by 30% for two such disturbances. Small changes in control limits can have large effects on the statistical performance of the chart. For time series with more than 200 observations, transfer function modeling is most important under conditions of large (greater than $4\sigma_a$) or multiple disturbances.

Our simulation studies indicate that point disturbances, confined to a single observation, are more easily detected than are step disturbances, with an effect persistent over many observations. It also appears to be easier to detect a given disturbance in an AR(1) time series than in an MA(1) time series, and detection rates increase for AR(1) models at higher values of the autoregressive parameter ϕ_1. The sensitivity of the transfer function model to detect outliers is dependent on the choice of the threshold, C, used in the iterative procedure. Operating characteristic curves constructed from the simulated results suggest that a threshold value of $C = 2.5$ provides a good tradeoff, resulting in high sensitivity while maintaining a false positive rate of less than 1%. For a threshold of 2.5, approximately 79% of $3\sigma_a$ disturbances are detected in AR(1) series and 66% in MA(1) series.

The need to analyze quality measurements for a time series structure is well documented in the quality literature. Perhaps less well understood are the benefits of checking process inputs for time series structure. If process inputs vary over time, they produce an additional source of variability that should be modeled. When analyzing a wastewater treatment plant, for example, we found that the input BOD had an MA(1) structure. By incorporating this information into the transfer function model for wastewater treatment, we were able to reduce our estimate of σ_a by 6.4%.

A disadvantage of the application of time series methods to process control is the loss of simplicity which is characteristic of Shewhart charts. We would like to emphasize that the computational burden of transfer function modeling is not as formidable as it may at first appear. The determination of ARIMA models with intervention terms and dynamic relationships is only performed periodically, when the identification of a new model is appropriate. Commercial software products are available that have automated routines for this analysis. Weller (1994) and Kusters (1995) identify PC-Expert from Scientific Computing Associates (www.scausa.com) and Autobox 3.0 (www.autobox.com) as two of the leading commercial products with capabilities for intervention analysis and transfer function modeling. The daily monitoring of the process can be accomplished by transposing the transfer function model into a spreadsheet format. For example, if Equation (11) were the basis of wastewater process monitoring, the operator would be required to enter only an input and output BOD rate. With this information, an appropriate monitoring strategy can be employed. For example, a residual can be determined for the residual chart of Alwan and Roberts (1988) or an intervention statistic calculated for the model of Atienza, Tang, and Ang (1998).

References

ALWAN, L. C. (1992). "Effects of Autocorrelation on Control Chart Performance". *Communications in Statistics: Theory and Methods* 21, pp. 1025–1049.

ALWAN, L. C. (2000). *Statistical Process Analysis* . Irwin McGraw Hill, New York, NY.

ALWAN, L. C. and RADSON, D. (1992). "Time-Series Investigation of Subsample Mean Charts". *IIE Transactions* 24, pp. 66–80.

ALWAN, L. C. and ROBERTS, H. V. (1988). "Time Series Modeling for Statistical Process Control". *Journal of Business and Economic Statistics* 6, pp. 87–95.

ALWAN, L. C. and ROBERTS, H. V. (1995). "The Problem with Misplaced Control Limits". *Applied Statistics* 44, pp. 269–278.

ATIENZA, O. O.; TANG, L. C.; and ANG, B. W. (1998). "A SPC Procedure for Detecting Level Shifts of Autocorrelated Processes". *Journal of Quality Technology* 30, pp. 340–351.

BERTHOUEX, P. M.; HUNTER, W. G.; and PALLESEN, L. (1978). "Monitoring Sewage Treatment Plants: Some Quality Control Aspects". *Journal of Quality Technology* 10, pp. 139–149.

BOX, G. E. P.; JENKINS, G. M.; and REINSEL, G. C. (1994). *Time Series Analysis, Forecasting and Control* . Prentice Hall, Englewood Cliffs, NJ.

BOX, G. E. P. and TIAO, G. C. (1975). "Intervention Analysis with Applications to Economic and Environmental Problems". *Journal of the American Statistical Association* 70, pp. 70–79.

BOYLES, R. A. (2000). "Phase I Analysis for Autocorrelated Processes". *Journal of Quality Technology* 32, pp. 395–409.

CHANG, I.; TIAO, G. C.; and CHEN, C. (1988). "Estimation of Time Series Parameters in the Presence of Outliers". *Technometrics* 30, pp. 193–204.

CHEN, C. and LIU, L. (1993a). "Joint Estimation of Model Parameters and Outlier Effects in Time Series". *Journal of the American Statistical Association* 88, pp. 284–297.

CHEN, C. and LIU, L. (1993b). "Forecasting Time Series with Outliers". *Journal of Forecasting* 12, pp. 13–35.

FOX, A. J. (1972). "Outliers in Time Series". *Journal of the Royal Statistical Society* B 43, pp. 350–363.

KIM, B. and MAY, G. S. (1994). "An Optimal Neural Network Model for Plasma Etching". *IEEE Transactions on Semiconductor Manufacturing* 7, pp. 12–21.

KUSTERS, U. (1995). "Software review: Autobox 3.0 and SCA PC-Expert. A review of two systems for automatic ARIMA forecasting". *The Forum* 8, pp. 1–6.

LU, C. W. and REYNOLDS, M. R., JR. (2001), "CUSUM Charts for Monitoring an Autocorrelated Process". *Journal of Quality Technology* 33, pp. 316–334.

MONTGOMERY, D. C. and MASTRANGELO, C. M. (1991). "Some Statistical Process Control Methods for Autocorrelated Data". *Journal of Quality Technology* 23, pp. 179–204.

MONTGOMERY, D. C. and JOHNSON, L. A. (1976). *Forecasting and Time Series Analysis* . McGraw Hill, New York, NY.

NOTOHARDJONO, B. D. and ERMER, D. S. (1986). "Time Series Control Charts for Correlated and Contaminated Data". *Journal of Engineering for Industry* 108, pp. 219–226.

POCH, M.; BEJAR, J.; and CORTES U. (1993). *Wastewater Treatment Data Source Information* . Universitat Autonoma de Barcelona and Universitat Politecnica de Catalunya, Barcelona, Spain.

SMITH, A. E. (1993). "Predicting Product Quality with Back-propagation: A Thermoplastic Injection Molding Case Study". *International Journal of Advanced Manufacturing Technology* 8, pp. 252–257.

WELLER, B. R. (1994). "Software Reviews, SCA PC-Expert for Windows". *International Journal of Forecasting* 10, pp. 481–487.

WEN, C. H. and VASSILIADIS, C. A. (1998). "Applying Hybrid Artificial Intelligence Techniques in Wastewater Treatment". *Engineering Applications of Artificial Intelligence* 11, pp. 685–705.

WOODALL, W. H. (2000). "Controversies and Contradictions in Statistical Process Control". *Journal of Quality Technology* 32, pp. 341–378.

WRIGHT, C. M.; BOOTH, D. E.; and HU, M. Y. (2001). "Joint Estimation: SPC Method for Short-Run Autocorrelated Data". *Journal of Quality Technology* 33, pp. 365–378.

XIA, Q.; RAO M.; SHAN, X.; and SHU, H. (1994). "Adaptive Control of a Paperboard Machine". *Pulp and Paper Canada* 95, pp. 51–55.

Key Words: *Autocorrelation, Statistical Process Control, Time Series Analysis.*