J. Chem. Inf. Comput. Sci. **1997,** *37,* 241−248

A Robust Smoothing Approach to Statistical Process Control

John Grznar and David E. Booth*

Department of Administrative Sciences, Graduate School of Management, College of Business Administration, Kent State University, Kent, Ohio 44242

Paul Sebastian

Evans School of Business, The University of Rio Grande, Rio Grande, Ohio 45353

Received May 8, 1996[⊗]

It has previously been shown that smoothing algorithms can provide the basis for methods to detect nuclear material losses and moreover can also provide a general approach to industrial statistical process control. The present paper extends this result by showing that a set of robust smoothers also produces methods that can be used in statistical process control. Further, it is shown that these smoothers are somewhat more sensitive to out of control points than those methods previously studied. The methods are successfully illustrated on chemical process data.

1. INTRODUCTION

Traditional control charts in use according to modern statistical process control (SPC) methodology, e.g. the $\bar{X}$ chart developed by Shewart,[36] assume the data to be independent and identically distributed (IID). However, this assumption is often violated in real situations. A common scenario in which the IID assumption does not hold is when the data is autocorrelated, i.e. high points tend to be followed by high points and low points tend to be followed by low points.[3,4,38] When the IID assumption does not hold, traditional control charts may be ineffective and inappropriate for monitoring and controlling product quality.[35,41] It is far more realistic, based on previous research, to assume that autocorrelations among the data exist.[1,2,34] It is the purpose of this paper to present a method for statistical process control that utilizes the fact that the data are autocorrelated. Thus, since the method would be based on a more accurate assumption, it should lead to superior results.

The method presented in this paper is to identify outliers in the process control data. Since outliers, by definition, are data that come from a different distribution than the main set of data, the presence of outliers would indicate the process being out of statistical control. This outlier detection is achieved by using robust smoothing techniques to smooth process control data. The smoothed data are then subtracted from the original data, resulting in an ordered set of residuals. The residuals can then be analyzed to identify the outliers by using the Outlier Detection Subroutine for Process Control (ODSPC) developed by Sebastian.[34]

It is assumed that users of these methods have at their disposal adequate computer power. Thus, a simpler, but equally accurate, method would not be of significantly greater value nor would a more complicated method (within reason) be a detriment.

2. LITERATURE REVIEW

The idea of using smoothing techniques to find outliers in statistical process control data is not new. Several authors have used various smoothing techniques with varying degrees of success.

Booth[4] and Booth et al.[5] used a time series method called the Generalized-M (GM) procedure, developed by Denby and Martin,[16] for the detection of changes in a process. The GM procedure provides information on the type of outlier (additive or innovative); however, the method is limited to those processes that can be modeled in terms of an autoregressive AR(p) process. This restriction limits the applicability of such a technique.[30] As we will discuss, knowing the outlier type helps to troubleshoot a process problem.

Utilizing the experience of Booth[4] in the use of the GM procedure, Prasad[30] used the M-type iterative procedure, developed by Chang,[9] Chang and Tiao,[10] Hillmer et al.,[25] and Chang et al.[11] to locate and identify outliers. Because it is based upon the Box−Jenkins[7] ARMA and ARIMA models, the procedure has wider applicability than the GM procedure.

Prasad[30] found that the method was able to detect outliers earlier than the traiditional control charts in data sets involving nuclear material accounting and industrial processes, because neither the standard 3σ rule nor the associated runs rules must be checked. Prasad also found that this outlier detection method is relatively robust to model misspecification, which is common if outliers are present. Once the outliers are identified, the model identification and parameter estimation can then be refined.[34]

Prasad et al.[31] have since utilized the joint estimation procedure of Chen and Liu[12] to improve upon the method developed in Prasad.[30]

Chen and Liu[12] developed the joint estimation procedure, which is more robust than the M-type Iterative Procedure. Huber[26] defines robustness as insensitivity to small deviations from assumptions. (Note: Bowen and Bennett[6] describe a point estimator as being robust if it is not seriously affected by departures from distributional assumptions or by potentially extreme sample observations.)

[⊗] Abstract published in *Advance ACS Abstracts,* February 15, 1997.

S0095-2338(96)00031-5 CCC: $14.00

The Prasad et al.[31] method can be used to identify not only additive outliers and innovative outliers, but also level shifts and temporary changes. If an outlier has properties of two types, then the dominant type will be identified. The importance of outlier type as we shall see is that it helps the operator determine assignable causes.

Sebastian's[34] methods use polynomial smoothing and data bounding techniques. Sebastian[34] also developed an outlier detection subroutine for process control (ODSPC), which identifies outliers by comparing the smoothed curve with the original curve. Basically, if the difference between the smoothed curve and the original curve (i.e. the residual) at point x_t is sufficiently large, then the point is deemed to be an outlier.

Hamburg[22] used a neural network approach developed by Denton.[17] Denton's program formulates a set of network weights using standard back-propagation training or nonlinear optimization algorithms and line searches in conjunction with backward error propagation. The algorithm requires various parameters to be established. These include input nodes, output nodes, hidden nodes, training, and testing tolerances. The neural network computer program provides various outputs such as the final weights of the training phase and the results of the training and testing phases.[22]

There are four criteria by which these current methods can be improved: (1) earlier detection of when the process is out of statistical control, (2) fewer type I errors—a type I error in this case is concluding that the process is out of statistical control when, in fact, it is not (a type I error is also called a false alarm), (3) fewer type II errors—a type II error in this case is not detecting when a process has gone out of control when, in fact, it has, and (4) ability to distinguish the loss type, i.e. the outlier type.

In this paper two methods of robust (i.e. outlier resistant) smoothers will be used. They are the locally weighted exponential scatterplot (LOWES) smoother, developed by Cleveland[14] and further revised by Hastie[24] and Härdle,[23] and the running median smoother 4253EH,t developed by Velleman[39] and Velleman and Hoaglin.[40]

The research hypothesis of this paper is that since these smoothers are more robust to outliers than previously used techniques, they should yield larger residuals and therefore should made outliers easier to detect.

Since Sebastian[34] compared his results to those obtained by Booth[4] and Prasad,[30] who in turn compared their results to traditional control techniques, the results from this study will be compared only to those obtained by Sebastian,[34] Prasad et al.,[31] and Hamburg.[22]

3. STATISTICAL PROCESS CONTROL CHARTS

Simply put, "a control chart is a graphical means for determining if the underlying distribution of some measurable variable seems to have undergone a shift".[29] The most commonly used types of control charts are the $\bar{X}$, R, and p charts. Other more sophisticated charts are not as widely used.[32]

There are but two basic ways of determining when a process is out-of-control: (1) by identification of an outlier in the time series data, since by definition, the outlier would have to come from a probability distribution different than the process's in-control distribution and (2) by forecasting the next unit of output to be out of specification (a very short term forecast!). Various rules exist for reading control charts to determine if the process is in statistical control. For example, for the $\bar{X}$ chart according to Juran,[27] a process is not in statistical control if

1. One point is outside the control limits (±3 standard deviations)

2. Two out of three successive points are at 2 standard deviations or beyond

3. Four out of five successive points are at 1 standard deviation or beyond

4. Eight successive points are on one side of the center line

Notice that the first of these so-called runs rules is used to detect an outlier, and rules 2−4 are used to anticipate (forecast) the process going out-of-control. It must be mentioned that these rules are rules of thumb. They cannot be used to determine the precise probability that the next unit of output will be within specifications. Furthermore, knowledge gained from experience with the particular process being monitored should be used to supercede these rules. It is, perhaps, for this reason that many authors in this area no longer list such rules in their work—for example Juran.[28]

Another problem with using these rules (assuming normally distributed IID data) is that they can lead to an unacceptably large number of false alarms. For example, consider a production schedule that calls for a process to be run for a period of just 2 weeks, 7 days a week, three shifts a day with quality control checks to be made hourly (a not uncommon situation, especially for processes for which it is very expensive to shut down and start back up). The probability that this method will result in *at least one* out-of-control signal due to rule 1 alone is nearly 60%; for rule 2, the probability is greater than 48%. For rule 3 it is more than 65%, and rule 4 alone would result in at least one out-of-control signal 74% of the time. Thus it should be expected that the process would be shut down due to a false alarm at least a couple of times during the 2 week run. The Appendix is a chart of these probabilities for various run lengths.

4. OUTLIERS DEFINED AND DESCRIBED

The definition of outlier is a highly subjective definition. No definition exists that can be used to determine precisely when an observation is or is not an outlier. For example, the eighth edition of the APICS dictionary defines an outlier as "a data point that differs significantly from other data for a similar phenomenon." This definition gives no mention as to the reason why the outlier should differ significantly from the other data, nor does it give any objective criteria for determining when a data point "differs significantly" from the other data.

Outlier Types. Fox[18] identified two types of outliers in relation to time series analysis which were later named by Denby and Martin[16] as the additive outlier (AO) and the innovative outlier (IO). An AO affects a single observation at the time of the disturbance. It is related to a one-time only effect. On the other hand, an IO affects the observation at the time of the disturbance as well as following, though not necessarily consecutive, observations. It is tied to a continuing process problem where some process variable has changed. An outlier in control chart type data indicates a change in the underlying production process.

Chen and Tiao[13] and Chen and Liu[12] identified two other types of outliers: the temporary change (TC) and the level shift (LS). A TC outlier is due to a process variable changing to a new level and then reverting back to its original level, for example, when a contaminated raw material is no longer used. It is a continuing problem of longer duration than that with an IO. The LS outlier is a sudden and more permanent change in the process. It is of much longer duration than the process changes associated with the IO and TC outliers. It is conjectured that early identification of an outlier and its type would decrease the cost of downtime since the technician would have additional information when trouble-shooting the process.

Booth[4] related outliers found in a process to represent observations out-of-control, indicating quality control problems. It is a basic premise of this paper that the methods of SPC can be improved upon by considering quality control data as time series data (which it is) and by using methods to detect outliers in time series data to determine whether or not the process is in-control.

5. SMOOTHING DEFINED AND DESCRIBED

Smoothing of a time series is a technique that is used in an attempt to filter out random noise and other irregularities. Smoothing is based on the assumption that the data point at time *t should not* deviate too markedly from the data points at times $t - 1$ and $t + 1$. When such a point is found, it is adjusted to be more similar to its neighboring points. Note that if smoothing is taken to an extreme, then the resultant smooth is nothing but a straight horizontal line.

When the smooth curve is subtracted from the original curve, the resultant curve is known as the residual curve. It is assumed that the process when in-control will generate a residual curve that does not deviate much from zero. Large residuals then indicate the presence of significant random noise or other irregularities affecting the process. That is, large residuals indicate outliers.

Locally Weighted Exponential Scatterplot (LOWES) Smoothing. LOWES is a method for smoothing scatterplot, (x_i, y_i), $i = 1, ..., n$, data. Since equally spaced time series data is but a special case of scatterplot data, it is also applicable for smoothing time series data. An advantage such a method has over methods that are designed specifically for equally spaced time series data, e.g. least squares moving polynomial fit smoothing (see Savitzky and Golay[33] or Steinier et al.[37]), is that it is also applicable for the (not too uncommon) case of missing data or unequally spaced time series data.

The method of LOWES uses two sets of weights in computing the smoothed value at x. Note, for LOWES x need not be equal to some x in the set of x of the original (x_i, y_i) data; however, for this discussion there will be no need to calculate a smoothed value at $x \neq x_i \ \forall\ i$. Therefore, throughout the remainder of this discussion this point will be assumed. Only data points within a k-nearest neighborhood (k-NN) about x are given positive, nonzero weights, where k is the number of data points in the nearest neighborhood. (k is usually expressed as a percentage of the number of data points in the original (x_i, y_i) data.) All data points outside k-NN are given a weight of zero; thus they are noninfluential on the smoothed values, $\hat{y}$.

One set of weights, w_i, is based on a weight function, W, that yields large weights if x_i is close to x and small weights if it is not close to x, where x is defined in equation 3.1. The weight function W must have the following properties:

1. $W(w) > 0$ for $|w| < 1$;
2. $W(-w) = W(w)$;
3. $W(w)$ is a nonincreasing function for $w \geq 0$;
4. $W(w) = 0$ for $|w| \geq 1$.

Property 1 is required because negative weights do not make sense. Property 2 is required because there is no good reason why points to the left of x should be weighted any differently than points to the right of x. Property 3 merely states that the farther points are from x the less weight they should be given. Property 4 is included for computational ease; it is used to guarantee that only points within a k-NN range are given positive, nonzero weights.

For W Cleveland uses the "tricube" function,

$$W(x) = (1 - |x|^3)^3, \quad \text{for } |x| < 1$$

$$W(x) = 0, \quad \text{for } |x| \geq 1$$

because according to Cleveland the tricube function enhances a χ^2 distributional approximation of an estimate of the error variance and should provide an adequate smooth in almost all situations. Therefore, for this research the same function will be used.

To calculate the weights, $w_i(x)$, first find h, the largest number among $|x_i - x|$ for $x_i \in \{k\text{-}NN\}$. Then

$$w_i(x) = W(h^{-1}(x_i - x)) \tag{3.1}$$

The second set of weights, δ_i, called robustness weights by Cleveland, is based on the size of the residuals. Large residuals result in small weights and small residuals result in large weights. For the robustness weight function, Cleveland uses

$$\delta_i = w_i B(e_i/6s)$$

where, s is the median of the residuals and B is defined as the "bisquare" function

$$B(x) = (1 - x^2)^2, \quad \text{for } |x| < 1$$

$$B(x) = 0, \quad \text{for } |x| \geq 1$$

Cleveland uses the bisquare function because other investigations, e.g. Gross,[20,21] have shown it to perform well for robust regression, i.e. it is well-known that it accommodates for outliers, and hence provides a robust smoother. The bisquare function will likewise be used in this research.

Now that how to calculate the weights has been described, the LOWES algorithm can be listed in a step by step fashion. To find the smoothed point $(x_i, \hat{y}_i)$ the following sequence of operations is performed:

1. Compute the set of weights $w_i(x_i)$.
2. Run a weighted least squares polynomial regression of degree d using the weights w_i for (x_i, y_i). Let $\hat{y}_i$ be the fitted value of the regression at x_i.
3. Find the median, s, of the absolute valued residuals $e_i = |y_i - \hat{y}_i|$.
4. Compute the set of robustness weights δ_i.

5. Find new $\hat{y}_i$ by running a weighted least squares polynomial regression of degree d using the robustness weights, δ_i, for $(x_i, \hat{y}_i)$.

6. Repeat steps 3–5 a total of t times or, if desired, until convergence.

It can be seen from the above that the use of the LOWES algorithm requires the selection of the parameters k, d, and t, in addition to W and B.

Choosing *k*, *d*, and *t*. The parameter k is what determines the amount of smoothing. The larger k is, the more smooth the resulting curve is. Therefore, k needs to be chosen on the basis of the properties of the data and the amount of smoothing desired.

The parameter d is the degree of the polynomial used in the locally weighted regression and is restricted to being a non-negative integer. As d is increased, the computational complexity of the regression quickly rises; however, the flexibility to reproduce patterns in the data is enhanced. If d is set at $d = 0$, then local constancy is assumed. Since the data may or may not exhibit autocorrelation, d should be set at $d > 0$. Setting $d = 1$ should provide a good balance between computation ease and flexibility.

Experimentation with LOWES has shown that LOWES tends to converge rather quickly. Two iterations, i.e. $t = 2$, should be adequate. Alternatively, a convergence criterion could be established, and the procedure could be iterated, however many number of times it takes to meet the criterion. However, this would seem to be needlessly cumbersome.

Running Median Smoothing (RMS). RMS is a smoothing technique for equally spaced or almost equally spaced time series (t_i, y_i) data with a significant amount of autocorrelation. It is based on the idea that autocorrelated data should not deviate too markedly from other nearby (sequentialwise) data points. Basically, the smoothed value, $\hat{y}_i$, is the median of the y_i values within a symmetric neighborhood (span) about t_i. The span can be any positive nonzero integer value, although a span of size 1 makes no practical sense. Spans larger than 7 are rarely used in practice.[40]

The shorter the span is, the gentler the smooth is. In other words, smooths created by shorter spans retain more characteristics of the original data; unfortunately, this includes outliers. The longer the span is, the more resistant the smooth is to outliers, again providing the robustness property, but the smooth loses more characteristics of the original data. This, for our purposes, will tend to increase the number of false alarms.

There is a difference in the approach to using even-span and odd-span running medians. When using an odd-span running median, the smoothed value $\hat{y}_i$ is naturally recorded at t_i because t_i is located at the center of the span. But, with even-span running medians there is no t_i located at the center of the span; therefore, with even-span running medians the smoothed value $\hat{y}_i$ is recorded at the middle of the gap between the two middle values of t_i in the span. The data can be easily recentered at the original t_i values by resmoothing the data with another even-span running median smoother. Usually, this recentering is done with a running median of span 2.

When using RMS, data near the ends of the sequence require special considerations. For data near the ends of the sequence, the maximum allowable size of the span is restricted by the requirement that the span be a symmetric neighborhood about t_i. Therefore, if an odd (even)-span is being used to smooth the data, then for data near the ends of the sequence the span is reduced to the largest odd (even)-span that permits a symmetric neighborhood about t_i. For example, if a span of 5 is being used to smooth the data, then the next-to-the-endpoints are smoothed using a span of 3. In all cases the extreme endpoints are either just copied or smoothed by using endpoint extrapolation.

The idea of endpoint extrapolation is to first estimate what the next value past the endpoint might have been by using the two smoothed values on the inward side of the endpoint. That is, $\hat{y}_0$ is estimated by using $\hat{y}_2$ and $\hat{y}_3$, and $\hat{y}_{n+1}$ is estimated by using $\hat{y}_{n-1}$ and $\hat{y}_{n-2}$. Standard linear extrapolation is used when estimating $\hat{y}_0$ and $\hat{y}_{n+1}$. Then, $\hat{y}_1 = \text{med}\{\hat{y}_0, y_1, \hat{y}_2\}$ and $\hat{y}_n = \text{med}\{\hat{y}_{n-1}, y_n, \hat{y}_{n+1}\}$.

Other advanced RMS techniques (discussed below) listed in Velleman and Hoaglin[40] include Hanning, resmoothing, reroughing, twicing, and splitting.

Advanced RMS Techniques. Odd-span running median and even-span with recentering running median smoothing make up what are called elementary running median smoothing techniques. Unfortunately, using only an elementary RMS technique usually results in either a data pattern that does not have enough of the characteristics of the original data sequence or a data pattern that is not smooth enough. Therefore, the purpose of the advanced RMS techniques is to enhance the smoothing while keeping the loss of the characteristics of the original data sequence to a minimum.

Resmoothing. One method of improving upon the elementary RMS techniques is to apply one of the elementary RMS techniques to the results of another elementary RMS technique. This process is known as *resmoothing*. For example, after using a running median of span 5, the smoothed sequence could then be resmoothed with a running median of span 3.

Since, as mentioned above, spans greater than 7 are rarely used in practice, a natural way to denote a resmoothing combination is to merely concatenate the span sizes. Such as, for the example mentioned above, where we first smoothed with a span of 5 and then resmoothed with a span of 3, this combination can be referred to as a 53 combination.

Sometimes we may wish to use the same span repeatedly until further resmoothing produces no further change in the smoothed sequence. Such a combination can be denoted putting an "R" after the span size. For example, 3R would denote smoothing with a span of 3 and then repeatedly resmoothing the resulting sequence with a span of 3 until no further changes are produced.

Hanning. This technique is named after Julius von Hann, who advocated its use.[40] Hanning is a procedure whereby a data point is replaced with a weighted average of a number of data points located in a neighborhood about the data point. One form of Hanning is to use three data points with weights of $^1/_4$, $^1/_2$, and $^1/_4$. For example, y_t might be replaced with

$$\hat{y}_t = 0.25y_{t-1} + 0.5y_t + 0.25y_{t+1}$$

Obviously, any combination of weights could be used; the only criterion is that they sum to one. However, the $^1/_4$, $^1/_2$, $^1/_4$ set of weights is the most commonly used set for Hanning. Since Hanning can have a drastic effect on outliers, it is generally used only after outliers have been smoothed away by other techniques. Hanning is denoted by the letter "H", thus 53H would indicate the sequence was (or is to be) first smoothed with a span of 5, then resmoothed with a span of 3, and finally Hanned. Notice, this notation does not indicate which set of weights is used for the Hanning step; this must be indicated elsewhere.

Reroughing and Twicing. Subtracting the smooth sequence from the original sequence, corresponding term by corresponding term, yields what is called the *rough*. The rough is, of course, another sequence and as such it too can be smoothed. *Reroughing* is the process of smoothing the rough and then adding the smooth of the rough to the smooth of the original sequence. The purpose of reroughing is to restore some of the characteristics of the original sequence to the smoothed sequence.

Reroughing is not denoted by a single letter in the name of a smoothing sequence of operations, instead ", reroughed with ..." is used. For example, 53, reroughed with 42 means first a span of 5 and then a span of 3 is used on the original sequence, then the smooth generated by 53 is subtracted from the original sequence to yield the rough. The rough is then operated on with a span of 4 recentered with a span of 2. This yields a smooth of the rough. This smooth is then added to the smooth that resulted from the 53 operation.

Quite often the same sequence that was used on the original sequence is also used for reroughing, this is called twicing and is denoted by "twice." Therefore, 53, twice means to apply 53 to the original data and then apply the same set of operations, 53, to the rough.

Splitting. Some combinations of smoothers, e.g. 3R, have a tendency to chop off peaks and valleys and to leave flat "mesas" and "dales" two points long. That is, y_t and y_{t+1} constitute a mesa if $y_t = y_{t+1}$, $y_{t-1} < y_t$, and $y_{t+2} < y_{t+1}$. Likewise, y_t and y_{t+1} form a dale if $y_t = y_{t+1}$, $y_{t-1} > y_t$, and $y_{t+2} > y_{t+1}$. These mesas and dales can be reshaped to restore some peakedness to the smooth by using a process known as *splitting*. Six data points are involved in the process of splitting, two points to the left of the flat spot, the two point flat spot, and two points to the right of the flat spot. If y_t and y_{t+1} is a mesa or a dale, then y_{t-2}, y_{t-1}, and y_t are used to recompute y_t, and y_{t+1}, y_{t+2}, and y_{t+3} are used to recompute y_{t+1}.

The mechanics of splitting are similar to those of endpoint smoothing as described above. That is, if y_t and y_{t+1} is a mesa or a dale, then splitting results in

$$\hat{y}_t = \text{med}\{y_{t-1}, y_t, 3y_{t-1} - 2y_{t-2}\}$$

and

$$\hat{y}_{t+1} = \text{med}\{3y_{t+2} - 2y_{t+3}, y_{t+1}, y_{t+2}\}$$

The notation for splitting is the letter S. If consecutive Ss appear, then the splitting, each time, is done after performing the same sequence of operations. For example, 3RSS means to repeatedly use span 3 until no further changes occur, then split, then repeat the 3R operation, then split again.

Notation for Smoothing of the Endpoints. No special notation is used to denote how the endpoints are handled when the endpoints are just copied onto the smooth. When the endpoints are smoothed via the method of endpoint extrapolation, as described above, then the letter "E" is used to make this designation. Endpoint smoothing is seldom done after every operation, usually it is done only once near the end of the sequence of operations, e.g. 4253EH, twice.

We will compare the LOWES smoother and the 4253EH,t smoothers to the Savitsky−Golay[33] based polynomial smoothers of Sebastian.[34] It should be noted that the ODSPC developed by Sebastian[34] was used with the robust smoothers considered here as well as with the polynomial smoother. The first author developed computer programs to implement these smoothers in combination with Sebastian's ODSPC. The programs are available from him.

6. OUTLIER DETECTION SUBROUTINE FOR PROCESS CONTROL

After the data has been smoothed, the residuals (rough) $e_i = y_i - \hat{y}_i$ will be analyzed to determine if the data contains outliers. To do this analysis the ODSPC as developed in Sebastian[34] will be used.

Collett and Lewis[15] confirmed that identifying outliers by visually examining the graphs of the data is highly subjective in nature. The demonstrated that the same analyst would often arrive at different conclusions when examining the same data presented in different formats. Therefore, what is needed is an objective method for identifying outliers. The ODSPC was developed to be such a method.

Use of the ODSPC is not dependent on the particular smoothing technique used; that is, any smoothing technique may be used with the ODSPC.

If the user does not know the standard deviation when the process is in-control (σ), then the program can be set to use the standard deviation of the data (s). Furthermore, if the user has reason to believe that the time series data contains outliers, then a better proxy for σ can be calculated by removing a number of extreme data points equal to the expected number of outliers before computing the process average and the process standard deviation. These two statistics are called the modified process average (*MPA*) and the modified standard deviation (*MSD*), respectively. The user can also set the program to use either a known process mean or the modified process average for the target value (*TV*).

The ODSPC labels data points as outliers according to two criteria, a primary operationalization condition and a secondary operationalization condition. Since the secondary operationalization condition is independent of any smoothing, it should not be used nor discussed here.

The primary operationalization condition is such that the y_i is identified as an outlier if (a) $y_i > TV + Z_4(MSD)$ and $y_i > TV$ and $y_i > \hat{y}_i + Z_2(MSD)$ or (b) $y_i < TV - Z_4(MSD)$ and $y_i < TV$ and $y_i < \hat{y}_i - Z_2(MSD)$. Both Z_2 and Z_4 may be set by the user of the subroutine. That is, if there is a marked deviation ($Z_2(MSD)$) between it and the adjusted or smoothed time series or model in the appropriate direction. In other words, if y_i is part of a peak above the center line (*TV*), the outlier criterion is $y_i > \hat{y}_i + Z_2(MSD)$. If y_i is part of a trough below *TV*, the outlier criterion is $y_i < \hat{y}_i - Z_2(MSD)$. Z_2 is chosen according to the desired sensitivity, where $Z_2 = 0.5$ is the usual starting point or default value. Other peaks below *TV* and troughs above *TV* are ignored since it is doubtful that such points are outliers within a process control context.

It may happen that y_i and other parts of the adjusted graph or model do not accurately represent the underlying process.

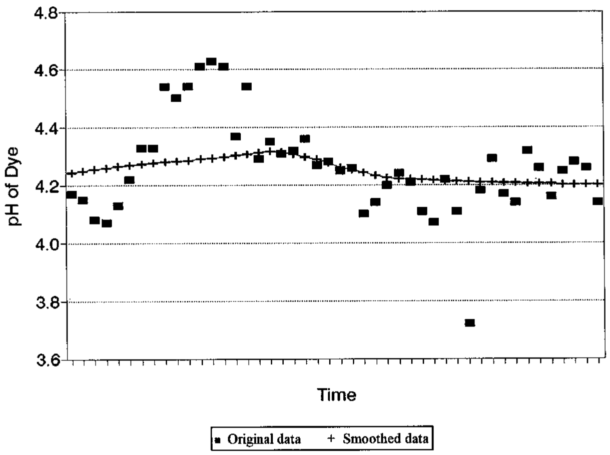

Figure 1. LOWES smoothing ($k = 0.8$). Sample averages of pH of dye liquor.

Then one may miss bona fide outliers (type II error) and falsely identify others (type I error). Obviously, no method can specify a model that gives an exact representation of the process. As George Box[8] has said, "All models are wrong, but some are useful."

7. COMPARISON AND ANALYSIS OF VARIOUS METHODS

For the empirical SPC data set, the dyeing of woolen yarns data set provided in Grant and Leavenworth[19] will be used. This data set was chosen because it contains both additive and innovative outliers. The data set consist of 46 data points, each one being the average of pH values across five Husong kettles over time. A low pH value corresponds to high acidity, and vice versa. The acidity of the dyeing solution depends not only on the constituents put into the dye liquor but also on the characteristics of the wool being dyed. From observation 9 through 14, the acidity dropped below in-control levels because a different blend of wools was being used. Since this out-of-control situation would persist as long as these different wools were being used, these outliers should be considered as continuing outliers (COs). That is, a CO can be either an IO, a TC, or a LS type of

Table 1. Detection of Outliers among Sample Averages of pH of Dye Liquor[a]

time period	joint estimation	data bounding	polynomial smoothing	neural network	robust smoothing	assigned cause
1						
2						
3				*	CO	different wools
4				*	CO	different wools
5						
6						
7						
8						
9	*	*	*	*	CO	different wools
10		*		*	CO	different wools
11		*	*	*	CO	different wools
12		*	*	*	CO	different wools
13		*	*	*	CO	different wools
14		*	*	*	CO	different wools
15						corrective measures
16	*	*	*	*	AO	corrective measures
17						corrective measures
18						
19						
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32				*		none/unknown
33						
34						
35	*	*	*	*	AO	improperly neutralized
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						
46						

[a] * = detection of outlier; one-time = AO; continuing = CO.

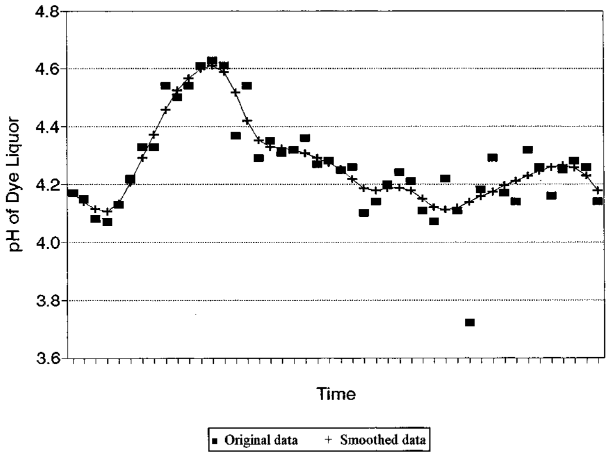

Figure 2. 4253EH, twice smoothing. Sample averages of pH of dye liquor.

outlier. During observations 15−17, corrective measures were being taken. Of these three observations, only number 16 was beyond the in-control limit. Since this was a temporary one-time event, it must properly be called on AO. Observation 35 is also an AO due to the use of improperly neutralized carbonized stock.

The low acidity indicated by observations 3 and 4 are also due to a different type of wool being used, and although the levels are slightly above the lower control limit, they should by definition be considered COs.

A graph of the LOWES smooth is shown in Figure 1 and a graph of the 4253EH,t smooth is shown in Figure 2 along with the original control chart data. It can be seen from these two figures that the LOWES smoother is more robust to COs than the 4253,t smoother. For example, notice how with the LOWES smooth observations 9−14 (COs) do not seem to "pull up" the smooth curve nearly as much as they do with the 4253,t smooth curve. As for AOs, notice the effect observation 35 has on the two smooths. Neither one seems to be affected much by its presence. Since the LOWES smoother is more robust to COs than the 4253EH,t smoother, and they are both robust to AOs, in conjunction they can be used to identify the outlier type. That is, if an outlier is detected by using the LOWES smoother, but not with the 4253EH,t smoother, then the outlier can be labeled as a CO. If an outlier is detected using both methods, then the outlier should be labeled as an AO. This is precisely what our results show. The results of our study are summarized in Table 1.

Observations 9−14 were detected as outliers by the LOWES method but not by the 4253EH,t method; thus they can be properly identified as COs. Observations 16 and 35 were detected by both methods and therefore are correctly labeled as COs. It is interesting to note that only our method and Hamburg's[22] neural network method were able to detect observations 3 and 4 as outliers. Also, our method resulted in no type I or type II errors.

We believe it is clear from this study that our method of outlier detection presents an improvement over previously published methods. The reason being, the LOWES and 4253EH,t smoothers are more robust, thus outliers protrude more, making them easier to detect. The reason that this approach is better than standard control charts is that our approach allows for tighter limits than a standard chart, and thus we do not need to wait as long as a control chart user

Table 2

runs length	rule 1	rule 2	rule 3	rule 4
1	0.0027	n/a	n/a	n/a
2	0.0054	0.0010	n/a	n/a
3	0.0081	0.0031	n/a	n/a
4	0.0108	0.0050	0.0013	n/a
5	0.0134	0.0070	0.0055	n/a
6	0.0161	0.0089	0.0091	n/a
7	0.0187	0.0109	0.0126	n/a
8	0.0214	0.0128	0.0160	0.0078
9	0.0240	0.0148	0.0194	0.0117
10	0.0267	0.0167	0.0228	0.0156
20	0.0526	0.0358	0.0561	0.0544
30	0.0779	0.0546	0.0882	0.0917
40	0.1025	0.0730	0.1191	0.1275
50	0.1264	0.0910	0.1488	0.1620
60	0.1497	0.1087	0.1773	0.1950
70	0.1724	0.1261	0.2048	0.2268
80	0.1945	0.1431	0.2313	0.2573
90	0.2160	0.1598	0.2568	0.2866
100	0.2369	0.1761	0.2813	0.3148
125	0.2868	0.2156	0.3386	0.3804
150	0.3334	0.2532	0.3906	0.4397
175	0.3770	0.2891	0.4379	0.4934
200	0.4177	0.3231	0.4808	0.5419
225	0.4557	0.3556	0.5198	0.5857
250	0.4913	0.3865	0.5553	0.6254
275	0.5246	0.4159	0.5874	0.6613
300	0.5556	0.4439	0.6167	0.6937
325	0.5847	0.4706	0.6432	0.7230
336	*0.5968*	*0.4819*	*0.6541*	*0.7350*
350	0.6118	0.4960	0.6673	0.7495
375	0.6372	0.5201	0.6892	0.7735
400	0.6609	0.5432	0.7091	0.7952
425	0.6831	0.5651	0.7272	0.8148
450	0.7038	0.5859	0.7436	0.8326
475	0.7231	0.6058	0.7586	0.8486
500	0.7412	0.6247	0.7721	0.8631
525	0.7581	0.6427	0.7844	0.8762
550	0.7740	0.6598	0.7956	0.8880
575	0.7887	0.6761	0.8057	0.8988
600	0.8025	0.6916	0.8150	0.9085
650	0.8275	0.7205	0.8310	0.9252
700	0.8493	0.7467	0.8441	0.9388
750	0.8684	0.7704	0.8550	0.9500
800	0.8850	0.7919	0.8640	0.9591
850	0.8996	0.8114	0.8714	0.9665
900	0.9123	0.8290	0.8775	0.9726
950	0.9233	0.8450	0.8826	0.9776
1000	0.9330	0.8595	0.8867	0.9817

to get either an outlier signal or a runs rule violation. Further, the smoothing takes autocorrelation into account, and thus we do not have the problem of the violation of the IID assumption required by standard control chart methodology.

ACKNOWLEDGMENT

The authors would like to thank the Editor and the reviewers for their most helpful comments and suggestions.

APPENDIX

Probabilities of a Type I Error for In-Control Process. Table 2 shows the probabilities of getting a type I error (i.e. a false alarm) by means of the four runs rules mentioned in the paper. The data were computed assuming the in-control data were normally distributed and two-tailed test were used.

REFERENCES AND NOTES

(1) Alwan, L. C.; Radson, D. Implementation Issues of Time-Series Based Statistical Process Control. *Prod. Oper. Manage.* **1995**, *4*(3), 263−76.

(2) Alwan, Layth C.; Roberts, Harry V. The Problem of Misplaced Control Limits (with discussions). *Appl. Stat.* **1995**, *44*(3), 269−78.
(3) Alwan, Layth C., and Harry V. Roberts (1988). "Time Series Modeling for Statistical Process Control." *J. Bus. Econ. Stat.* **1988**, *6*(1), 87−95.
(4) Booth, David E. Some Applications of Robust Statistical Methods to Analytical Chemistry. Unpublished Ph.D. dissertation, University of North Carolina at Chapel Hill, 1984.
(5) Booth, David E.; Acar, William; Isenhour, Thomas L.; Ahkam, Sharif. Robust Times Series Models and Statistical Process Control. *Ind. Math.* **1990**, *40* (Part 1), 73−97.
(6) Bowen, W. M.; Bennett, C. A. *Statistical Methods for Nuclear Material Management*. U.S. Government Printing Office: Washington, DC, 1988.
(7) Box, G. E. P.; Jenkins, G. M. *Time Series Analysis Forecasting and Control*, 2nd ed.; Holden-Day: San Francisco, 1976.
(8) Box, G. E. P. Robustness in the Strategy of Scientific Model Building. In *Robustness in Statistics*; Launer, R. L., Wilkinson, G. N., Eds. Academic Press: New York, 1979; pp 201−236.
(9) Chang, I. Outliers in Time Series. Unpublished Ph.D. dissertation, University of Wisconsin at Madison, 1982.
(10) Chang, I.; Tiao, G. C. (1983). Estimation of Time Series Parameters in the Presence of Outliers. University of Chicago Statistical Research Center, Technical Report 8.
(11) Chang, I.; Tiao, G. C.; Chen, C. Estimation of Time Series Parameters in the Presence of Outliers. *Technometrics* **1988**, *30*(2), 193−204.
(12) Chen, Chung; Liu, Lon-Miu. Joint Estimation of Model Parameters and Outlier Effects in Time Series. *J. Am. Stat. Assoc.* **1993**, *88*, 284−297.
(13) Chen, C.; Tiao, G. C. Random Level Shift Time Series Models, ARIMA Approximation and Level Shift Detection. *J. Bus. Econ. Stat.* **1990**, *8*, 170−186.
(14) Cleveland, W. S. Robust Locally Weighted Regression and Smoothing Scatter Plots. *J. Am. Stat. Assoc.* **1979**, *74*, 829−836.
(15) Collet, D.; Lewis, T. The Subjective Nature of Outlier Rejection Procedures. *Appl. Stat.* **1976**, *25*, 228−237.
(16) Denby, L.; Martin, R. D. Estimation of the First Order Autoregressive Parameter. *J. Am. Stat. Assoc.* **1979**, *74*, 140−146.
(17) Denton, J. W. *Instruction Manual Optinet Neural Network Training Software*." Kent State University Press: Kent, OH, 1993.
(18) Fox, A. J. Outliers in Time Series. *J. R. Stat. Soc., Ser. B* **1972**, *34*, 340−63.
(19) Grant, E. L.; Leavenworth, R. S. *Statistical Quality Control*, 5th ed.; McGraw-Hill: New York: 1980.
(20) Gross, Alan M. Confidence Interval Robustness With Long-Tailed symmetric distributions. *J. Am. Stat. Assoc.* **1976**, *71*, 409−416.
(21) Gross, Alan M. Confidence Intervals for Bisquare Regression Estimates. *J. Am. Stat. Assoc.* **1977**, *72*, 341−354.
(22) Hamburg, J. H. The Application of Neural Networks to Production Process Control. Unpublished Ph.D. dissertation, Kent State University, 1996.
(23) Härdle, W. *Applied Non-Parametric Regression*; Cambridge University Press: New York, 1990.
(24) Hastie, T. J.; Tibshirani, R. J. *Generalized Additive Models*; Chapman and Hall: New York, 1990.
(25) Hillmer, S. C.; Bell, W. R.; Tiao, G. C. Modeling Considerations in the Seasonal Adjustment of Economic Time Series. In *Applied Time Series Analysis of Economic Data*; Zellner, A., Ed.; U.S. Bureau of the Census: Washington, 1983.
(26) Huber, P. J. *Robust Statistics*; Wiley: New York, 1981.
(27) Juran, J. M. *Juran's Quality Control Handbook*, 3rd ed.; McGraw-Hill: New York, 1974.
(28) Juran, J. M. *Juran's Quality Control Handbook*, 4th ed.; McGraw-Hill: New York, 1988.
(29) Nahmais, Steven *Production and Operations Analysis*, 2nd ed.; Irwin: Boston, 1993.
(30) Prasad, Sameer. A Robust Monitoring Method for the Early Detection of Deviations in a Time Series Process with Applications in Operations Management. Unpublished Ph.D. dissertation, Kent State University, Kent, OH, 1990.
(31) Prasad, Sameer; Booth, David; Hu, Michael; Deligonul, Seyda. The Detection of Nuclear Materials Losses. *Decis. Sci.* **1995**, *26*(2), 265−81.
(32) Saniga, Erwin M.; Shirland, Larry E. Quality Control in Practice—A survey. *Qual. Prog.* **1977**, *10*(5), 30−33.
(33) Savitzky, Abraham; Golay, Marcel J. E. Smoothing and Differentiation of Data By Simplified Least Squares Procedures. *Anal. Chem.* **1964**, *36*, 1627−1639.
(34) Sebastian, Paul R. Polynomial Smoothing and Data Bounding Applied to Industrial Process Control and Nuclear Material Safeguards. Unpublished Ph.D. dissertation, Kent State University, Kent, OH, 1994.
(35) Sebastian, Paul R.; Booth, David E.; Hu, Michael Y. Using Polynomial Smoothing and Data Bounding for the Detection of Nuclear Material Diversions and Losses. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 442−450.
(36) Shewhart, W. A. *Economic Control of the Quality of Manufacturing*; Van Nostrand: New York, 1931.
(37) Steinier, Jean; Termonia, Yves; Deltour, Jules. Comments on Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* **1972**, *44*, 1906−1909.
(38) Vasilopaulous, A. V.; Stamboulis, A. P. Modification of Control Chart Limits in the Presence of Data Correlation. *J. Qual. Technol.* **1978**, *10*(1), 20−30.
(39) Velleman, Paul F. Definition and Comparison of Robust Nonlinear Data Smoothing Algorithms. *J. Am. Stat. Assoc.* **1980**, *75*, 609−615.
(40) Velleman, Paul F.; Hoaglin, David C. *Applications, Basics, and Computing of Exploratory Data Analysis*; Duxbury Press: Boston, 1981.
(41) Wardell, Don G.; Moskowitz, Herbert; Plante, Robert D. Control Charts in the Presence of Data Correlation. *Manage. Sci.* **1992**, *38*(8), 1084−1105.

CI9600317