The Use of Robust Smoothers in Nuclear Material Safeguards

John Grznar and David 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 April 25, 1996[⊗]

It has previously been shown that smoothing algorithms can provide the basis for a method to detect nuclear material diversions and 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 equivalent methods that can be used in nuclear material safeguards algorithms. Further, it is shown that these smoothers are somewhat more sensitive to loss points than the previously studied smoothers. The method is illustrated on real data.

INTRODUCTION

In nuclear materials accounting, the goal is to accurately discover materials losses (e.g. leakage or theft) as early as possible in order to minimize the threat to life and property. The theft of nuclear material is a serious problem in today's world. For example, German authorities reported that the number of nuclear-smuggling incidents has increased over the past few years. They report 41 incidents in 1991, 158 in 1992, 241 in 1993, and 267 in 1994.[1,2] Such results are clearly unacceptable. The purpose of this paper is to develop statistical tools so that such incidents can be stopped.

As has been shown previously,[3] nuclear material loss situations may be recognized through the detection of statistical outliers. In general terms, an outlier is an inconsistent or nonrepresentative point in a data set that causes a difficulty or problem in a statistical analysis. Traditional statistical process control (SPC) charts as well as most of the other methods in current use are based upon the assumption that the system observations in the time series are independent and identically distributed (IID) about the targeted process mean (μ_o or TV) at any time t and that the distribution is normal when the process is in statistical control.[3] Independence implies that there is no particular pattern in the data.

Points outside of three standard deviations of the targeted process mean (μ_o or TV) are usually considered to be outliers. If such exist, the process is said to be "out of control". In other words, there is a significant adverse process change due to an assignable cause. Otherwise the process is said to be "in control", i.e. random variations of output are within certain control limits.

Controlling the Process by Utilizing Time Series Models. In reality, the IID assumption of the conventional control chart often does not hold. In-control production process measures over time are often interdependent, i.e. the points are autocorrelated.[4−6] Under such conditions, traditional SPC procedures may be ineffective and inappropriate for monitoring and controlling product quality[3,17]—erroneously indicating an out-of-control situation when the supplementary criteria of the traditional control chart are applied. As we shall see, this is also true in nuclear material safeguards.

Many production process time series exhibit a characteristically repetitive pattern which can be mathematically modeled by an autoregressive moving average [ARMA(p,q)] model.[6] For example, ARMA(1,1) and other time series models have been empirically found in some cases to be appropriate for modeling a process time series.[2,3,7,8,17] Then an outlier, any point that deviates significantly from the underlying process model or time series pattern, indicates an out-of-control situation with respect to the process model. Such a point can be identified using statistical methods.

APPLICATIONS TO NUCLEAR MATERIAL SAFEGUARDS AND OTHER PRIOR RESEARCH

Statistical process control methodology is also applicable to other process industries such as the control of inventories of materials which are particularly valuable or involve critical safety concerns. Nuclear material safeguards are methods based upon nuclear material accounting to detect and prevent losses.

Common statistical problems in material balance accounting are treated by Goldman et al.[7] in detail. These include additive errors that often occur in sums of material inventory measurements and transfers which usually are estimated. Estimates of material held up in equipment (in-process) are especially problematic. Speed and Culpin[9] note that it is extremely difficult to differentiate between a systematic error over several periods such as a poorly calibrated instrument (a discrepancy) and an outright diversion or theft. Statistical methods that are in current use today have been roundly criticized by numerous authors.[3,7,8]

Material Balance Equation. The nuclear material balance at time t is defined as[9]

$$MB_t = I_t - I_{t-1} - T \quad (1)$$

where MB_t = observed material balance at time t (loss = −); I_t = inventory or quantity in the specified area at time t, including stock and material in-process; and T = net transfer acccounted for (in = +, out = −) during the time

⊗ Abstract published in *Advance ACS Abstracts,* February 1, 1997.

interval. If T is positive, in-process and stock increase. If there are no losses between time $t - 1$ and t, then $MB_t = 0$ and $I_t - I_{t-1} = T$, i.e. the net material transferred in (+) or out (−) to account for the change in inventory.

Another term, often used in the nuclear industry instead of material balance, is material unaccounted for (MUF_t). To use MUF_t, both sides of the above equation are multiplied by −1 and $-MB = MUF$, being thus positive for a loss.[11] Equation 1 then becomes

$$MUF_t = I_{t-1} - I_t + T \qquad (2)$$

or

$$MUF_{t+1} = I_t - I_{t+1} + T$$

A raw material inventory (of quantity I_t) consists of stock and the amount in-process. Transfer material (T) consists of new stock (+) and the raw material converted to waste and/or the final product leaving the process (−). It can be seen in eq 2 that if I_t is underestimated, $MUF_t > 0$. However, in the next period, $MUF_{t+1} < 0$ if I_{t+1} is estimated correctly.[11] On the other hand, if there is an actual loss at time t only and no measurement error, then $MUF_t > 0$. But $MUF_{t+1} = 0$ since all subsequent inventories will reflect that one time loss.

A positive MUF_t (negative MB_t) would be due to such causes as leakage, unanticipated loss in effluents (waste), theft, accounting errors, systematic measurement errors (perhaps caused by poor instrument calibration), plugged probes, solid buildup in the process (e.g. settling in tanks), operator errors as misreading, mistranscribing, miscalculating, etc. What remains after all possible explanations are exhausted is the usual random error.[11]

A negative MUF_t (positive MB_t), an apparent gain, may be due to systematic measurement errors or an unaccounted for abnormal hidden amount in-process that later reversed itself as for example, caking on tube walls being released in some manner.

Since many estimates of the material are made in many parts of the process and since there are numerous sources of error, often cumulative, seldom is $MUF_t = MB_t = 0$, the target value. Then a crucial question arises, "Is a negative material balance due to normal statistical variations, a large error, or has there been a loss or diversion of nuclear material?"

Material Balances As a Time Series. Even if no losses or gross errors have occurred, there will be variations among successive *MB* or *MUF* measurements about the process mean. A succession of such material balance measurements forms a time series. As with the control of a product specification in a production process, the output (*MB* or *MUF* in this case) may be monitored for outliers which are defined as significant deviations from the target value.

A history of methods to deal with the time series aspect of this problem can be found in the literature.[3]

PROPOSED ALGORITHMS FOR THE DETECTION OF OUTLIERS BASED ON ROBUST SMOOTHERS

The present paper studies three methods of robust (i.e. outlier resistant) smoothers to deal with the nuclear safeguards problem. We consider in order the LOWES smoother, developed by Cleveland,[10−12] and two running median smoothers 4253EH and 3RSSH developed by Velleman.[13,14] These three techniques were chosen because of their high resistance (robustness) to outliers; thus they should yield larger residuals than less resistant methods, resulting in a more correct placement of the smoothed curve. Furthermore, larger residuals should make outliers easier to detect.

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,[19,20] is that it is also applicable for the (not too uncommon) case of missing data or unequally spaced time series data.

Running median smoothing (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. Because the LOWES and the RMS smoothers are somewhat complex, we refer the reader to Cleveland[10] (for LOWES) or Velleman and Hoaglin[14] (for RMS) for the full mathematical details.

We will compare the LOWES smoother and the 4253EH and 3RSSH RMS smoothers to the Savitsky-Golay based polynomial smoothers of Sebastian et al.[3,15] It should be noted that the outlier detection subroutine for process control (ODSPC), developed by Sebastian et al.,[3,15] 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.

Since the cost of overlooking an outlier is very high in the case of nuclear materials, management may tolerate a certain frequency of false alarms. At the very least, a questionable outlier should serve as an early warning that a loss may be occurring. Management should be alert, and the operator should perform at least a cursory examination of the critical points in the process, look for a trend, and then monitor the process more closely.

TESTING THE ALGORITHMS ON DATA SETS WITH KNOWN DIVERSIONS

To obtain a more definitive test, we used nuclear material balance data sets with known outliers (losses), i.e. actual out-of-control nuclear inventories due to known protracted removals of nuclear material.

A series of experiments were conducted in 1980 at the Allied-General Nuclear Services (AGNS) Barnwell Nuclear Fuels Plant in order to improve the detection of nuclear material losses in the different unit process accounting areas (UPAAs). The material balance data were recorded periodically on a near real-time basis with the aid of a computerized nuclear materials control and accounting system.[9]

The experiments essentially involved physically removing nuclear material in different forms from a plutonium purification process (PPP) with variations in what was included in the UPAA for the material balance measure, location of removals, time, quantity, concentration, and type

Table 1. Location and Timing of the Removals for Four of the Minirun Experiments

experiment and location	tank	start time/date	end time/date	total removed obs	total removed vol (L)	U (kg)
minirun 3C		0815/07/18	1204/07/21			
	2AP	1645/07/18	0845/07/19	8−24	96	4.224
UPAA PPP	3AP	1645/07/18	1645/07/19	8−32	96	4.253
	1BP	1615/07/20	0850/07/21	56−72	96	5.722
U settled on bottom of acid concentrators = 120 kg						
minirun 4A		0700/09/04	1900/09/07			
UPAA 1BP	1BP	0615/09/06	0545/09/07	48−71	299.4	17.64
minirun 4B		0700/09/04	1200/09/08			
	3BP	1141/09/04	0445/09/06	4−45	370.3	20.99
UPAA column & sample Tk	1BP	0615/09/06	0545/09/07	47−70	299.4	17.64
	2BP	0900/09/07	1045/09/08	74−99	826	22.1
U settled on bottom of acid concentrators = 100 kg						
minirun 5B		1600/11/18	1630/11/21			
	1BP	0000/11/19	1200/11/19	9−20	174	10.2
UPAA column	1BP	1445/11/19	0115/11/20	23−34	177.2	11.1

Table 2. Summary of Minirun 3C Results[a]

	PS	R10	R20	R30	R40	R50	R60	R70	R80	R90	RM/A	RM/B
TOI	12	13	15	18	20	22	22	22	22	22	22	22
T + OI	7	7	8	11	13	15	15	15	15	15	15	15
NA + I	4	4	4	5	6	6	6	6	6	6	6	6
N + FA	3	3	4	6	7	9	9	9	9	9	9	9
N + OM	22	22	22	21	20	20	20	20	20	20	20	20

[a] The notation used in the summary table is as follows: *TOI* = number of observations identified as outliers. *T + OI* = number of positive observations identified as outliers. *NA + I* = number of positive observations during removal periods identified as outliers. *N + FA* = number of positive observations during nonremoval periods identified as outliers (i.e. false alarms). *N + OM* = number of positive observations during removal periods not identified as outliers. PS = polynomial smoothing. R*xx* = *xx*% of the observations were included in the neighborhood for LOWES. RM/A = running median smoother using 4253EH, twice. RM/B = running median smoother using 3RSSH, twice. [b] ODSPC settings: %drop = 5%; $z_2 = 0.7$; $z_3 = 3$; $z_4 = 1.4$.

of diversion. The process includes pulse columns (2AP and 3AP), a surge tank (1BP), and strip columns (2BP and 3BP). The original researchers called these experiments "miniruns".

Table 1[16] gives a summary of the location and timing of removals for the minirun data sets. In these particular experiments, solutions of uranium separated from the plutonium are gradually removed from different tanks (i.e. protracted removals or trickle losses), and material balances are taken every hour across the corresponding UPAA.

Ideally, trickle losses should be detected in the observations immediately following the starting time and continuing through the observation immediately following the ending time according to column 5, "obs". The total protracted removal across the extended time period by volume of solution and the corresponding uranium content is recorded in columns 6 and 7.

During the protracted removal periods, it is doubtful that a negative *MUF* can indicate a removal or loss. On the contrary, a negative *MUF* should indicate a release of material held up somewhere in the process as caking in the tubes or precipitated solids returning to solution. Of course, measurement errors are common. If the preceding inventory measure (I_{t-1}) is erroneously low (see eq 2), MUF_{t-1} will be too high. However, if I_t is accurate, then the next observation (MUF_t) will be too low. We now consider particular miniruns.

Minirun 3C. Table 1 shows that there are three protracted removals from three tanks in this experiment, two of which are in part simultaneous: 0.4186 kg/h during periods 8−24, 0.1701 kg/h during periods 25−32, and 0.3366 kg/h during periods 56−72. The 120 kg that settled out in the concentrators averages to 1.58 kg/h. Thus the modified process average (*MPA* = +1.145 kg) and median = 1.65 kg are above the target value (*TV* = 0), but they should be even higher because of those known removals and precipitation. Over the entire period, the separated solids (120 kg) and the amount removed (14.199 kg) or 134.2 kg total should given an average *MUF* = 1.77 kg/h over the 76 periods.

Polynomial smoothing detected a number of losses.[3,17] The summary in column 1 of Table 2 shows that

TOI = total outliers identified = 12

T + OI = total positive outliers identified = 7

NA + I = number of actual positive outliers identified = 4

NFA = number of false alarms = 3

N + OM = number of positive outliers missed = 22

For an explanation of the results, consider the polynomial smoothing (PS) method. The PS method resulted in a *T + OI* = 7 and a *NA + I* = 4 which is 15.8% of the positive observations during the protracted removal periods. The difference

$$T + OI - NA + I = N + FA \tag{3}$$

or

$$7 - 4 = 3$$

the number of false alarms, which is 14.3% of the positive observations during the nonremoval periods. The number of known positive outliers during the protracted removal periods (26) less the number of actual positive outliers identified (*NA + I* = 4) gives the number of positive outliers

Table 3. Summary of Minirun 4A Results[a]

	PS	R10	R20	R30	R40	R50	R60	R70	R80	R90	RM/A	RM/B
TOI	23	24	27	27	28	29	29	29	31	31	31	31
$T+OI$	7	8	9	9	9	9	9	9	9	9	9	9
$NA+I$	3	4	5	5	5	5	5	5	5	5	5	5
$N+FA$	4	4	4	4	4	4	4	4	4	4	4	4
$N+OM$	7	6	5	5	5	5	5	5	5	5	5	5

[a] See Table 2 for a summary of notation. ODSPC settings: %drop = 7%; $z_2 = 0.4$; $z_3 = 3$; $z_4 = 0.9$.

Table 4. Summary of Minirun 4B Results[a]

	PS	R10	R20	R30	R40	R50	R60	R70	R80	R90	RM/A	RM/B
TOI	24	23	30	40	36	38	40	40	39	39	57	57
$T+OI$	20	19	26	36	32	34	36	36	35	35	53	53
$NA+I$	20	19	26	36	32	34	36	36	35	35	53	53
$N+FA$	0	0	0	0	0	0	0	0	0	0	0	0
$N+OM$	61	62	55	45	49	47	45	45	46	46	28	28

[a] See Table 2 for a summary of motation. ODSPC settings: %drop = 8%; $z_2 = 0.5$; $z_3 = 3$; $z_4 = 1$.

Table 5. Summary of Minirun 5B Results[a]

	PS	R10	R20	R30	R40	R50	R60	R70	R80	R90	RM/A	RM/B
TOI	5	9	9	10	12	14	15	14	14	14	19	19
$T+OI$	2	3	5	6	7	8	9	8	8	8	12	12
$NA+I$	2	2	3	3	3	3	3	3	3	3	3	3
$N+FA$	0	6	2	3	4	5	6	5	5	5	9	9
$N+OM$	7	6	6	6	6	6	6	6	6	6	6	6

[a] See Table 2 for summary of motations. ODSPC settings: %drop = 2%; $z_2 = 0.5$; $z_3 = 3$; $z_4 = 1$.

missed ($N + OM = 22$). The references[3,17] give full details for the polynomial smoothing results.

It is crucial to detect an outlier as early as possible within the protracted period of removals. Polynomial smoothing detected the very first observation (no. 8) in the initial protracted removal period. From the summary results on minirun 3C (Table 2), we can see that the robust smoothers were also successful.

In fact, LOWES R10 provides the same summary statistics as did polynomial smoothing. As R*xx* increases, we notice an increase in real outliers identified but also we see a corresponding increase in false alarms. Thus, it seems that LOWES with R*xx* between R10 and R40 is a reasonable choice to use in a nuclear safeguards system. These results are also supported in Tables 3−5. The EDA smoothers RM/A and RM/B also show good results in Table 2. They both detect more actual outliers and slightly decrease the number of false alarms.

Minirun 4A. The main thing to notice in this summary (Table 3) is that the results are consistant with those of minirun 3C. This is true both for the LOWES smoothers as well as for RM/A and RM/B. These results indicate that the methods are working well as safeguards techniques.

Minirun 4B. The results in this data set (Table 4) are particularly encouraging for the running median smoothers, RM/A and RM/B. The summary statistics in Table 4 show that they find more real outliers and provide a substantial decrease in the number of false alarms.

Minirun 5B. The results shown in Table 5 are similarly encouraging as were the results obtained for the other minirun data.

Analysis. The outliers in these data sets are very difficult to identify because each removal per period is so small and generally does not exceed the normal in-control chance fluctuations.

The percentages of positive *MUF* observations during the removal and nonremoval periods are quite similar, making it very difficult to detect losses. In minirun 3C (Table 2), for example, there is often little difference between the removal and nonremoval periods, i.e. 61.9% of the observations were positive during the removal periods vs 61.8% during the nonremoval periods. In miniruns 4A and 5B, there is a greater frequency of positive *MUFs* during the nonremoval periods.

On the basis of the difficulty factors for these data sets, we are particularly pleased with the results that show the robust smoothers, LOWES and RM/A, and RM/B, are successful for use in safeguards applications.

Precipitated Solids. In miniruns 3C and 4B, outlier detection is further complicated by the fact that 120 and 100 kg, respectively, of uranium solids settled to the bottom of the concentrators. If there is a steady state or equilibrium, that would be no problem, but new material is constantly entering the process. Since these solids apparently were not considered in the nuclear inventories, positive *MUFs* or losses would be recorded.

We must assume that the original researchers conducted the experiments properly and removed the solids from the concentrators before the beginning of each run, but there is no way of knowing at what rate the particles settled. If constant, the uranium precipitated out of the solution at a rate of 1.58 kg/h or period in minirun 3C and 1.0 kg/h in minirun 4B. This greatly exceeds the rate of protracted removals which range from 0.170 to 0.735 kg/h.

Thus there may be some confusion as to whether an identified outlier represents a bona fide loss (removal) or simply solids that precipitated out. Therefore, a number of false alarms might be expected. At the same time, however, the methods should be able to adapt to the settled out solids

Table 6. Expected Average *MUF* vs *MPA* (kg)

experiment	settled solids	av settled	total removed	av removed	expected av *MUF*	*MPA*
minirun 3C	120	1.58/h	14.199	0.187/h	1.77	1.145
minirun 4A			17.64	0.210	0.210	−0.278
minirun 4B	100	1.00	60.73	0.607	1.61	1.08
minirun 5B			21.3	0.313	0.313	0.040

if the rate of precipitation is relatively constant as part of the process.

Table 6 compiles the average solids settling per hour and the average amount of material removed per hour over the duration of each experiment. The sum of these gives an expected average *MUF*.

It can be seen that the modified process average (*MPA*) is consistently less than the expected average *MUF*, leading us to suspect a systematic measurement error. Only in minirun 4A is there a known measurement problem, i.e. the plugged probes.[17]

Summary. All of the methods had difficulties in detecting the known outliers because of the very small removals and other problems that were discussed. In comparison to the other methods,[3] the smoothers described in this paper did quite well.

DISCUSSION

Real-Time Process Control. We have seen that the methods used in this paper are effective across different applications but simple in concept, flexible, and adaptable to the process environment. Thus they may be more acceptable to practitioners than the more sophisticated methods. The approach is clearly useful in a variety of applications, particularly in the case of autocorrelated data and most likely to severely nonnormal distribution situations, often due to the presence of many outliers. The exact conditions are being determined in current research. The proposed methods adequately handle the end points, thus being able to detect outliers as soon as the latest data become available. For example, when obs 72 in minirun 3C was used as the last observation, all of the methods proposed in this paper correctly detected the end point as an outlier, thus indicating these methods can be used for control in real-time.

The most important feature of these methods, however, is the ability to identify outliers, i.e. adverse process changes, earlier than traditional methods (e.g. control charts), well within the usual three standard deviations from the target value. Thus we propose that these methods are a useful addition to those already available to operating managers involved in nuclear material safeguards.

REFERENCES AND NOTES

(1) Williams, Phil; Woessner. The Real Threat of Nuclear Smuggling. *Sci. Am.* **1996**, *274* (1), 40−46.
(2) Zimmermann, Tim; Cooperman, Alan The Russian Connection. *U.S. News World Rep.* **1995**, *119* (16), 56−68.
(3) Sebastian, P. R.; Booth, D. E.; Hu, Michael Y. Using Polynomial Smoothing and Data Bounding for the Detection of Nuclear Materials Divisions and Losses. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 442−450.
(4) Booth, D. E. Some Applications of Robust Statistical Methods to Analytical Chemistry. Ph.D. Dissertation, University of North Carolina at Chapel Hill, 1984.
(5) Wardell, D. G., Moskowitz, H.; Plante, R. D. Control Charts in the Presence of Data Correlation. *Manage. Sci.* **1992**, *38*, 1084−1105.
(6) Vasilopoulous, A. V.; Stamboulis, A. P. Modification of Control Chart Limits in the Presence of Data Correlation. *J. Qual. Technol.* **1978**, *10* (10), 20−30.
(7) Goldman, A. S.; Picard, R. R.; Shipley, J. P. Statistical Methods for Nuclear Material Safeguards: An Overview. *Technometrics* **1982**, *24*, 267−275.
(8) Speed, T. P.; Culpin, D. The Role of Statistics in Nuclear Materials Accounting: Issues and Problems. *J. R. Stat. Soc., Ser. A* **1986**, *149*, Part 4, 281−313.
(9) Dayem, H. A.; Baker, A. L.; Cobb, D. D.; Hakkila, E. A.; Ostenak, C. A. *Demonstrations of Near-Real-Time Accounting: The AGNS 1980−1981 Miniruns*; Los Alamos National Laboratory: Los Alamos, NM.
(10) Cleveland, W. S. Robust Locally-Weighted Regression and Smoothing Scatterplots. *J. Am. Stat. Assoc.* **1979**, *74*, 829−36.
(11) Hastie, T. J.; Tibshirani, R. J. *Generalized Additive Models*; Chapman and Hall: New York, 1990.
(12) Härdle, W. *Applied Nonparametric Regression*; Cambridge University Press: Cambridge, England, 1990.
(13) Velleman, P. F. Definition and Comparison of Robust Nonlinear Data Smoothing Algorithms. *J. Am. Stat. Assoc.* **1980**, *75*, 609−15.
(14) Velleman, P. F.; Hoaglin, D. C. *Applications Basics and Computing of Exploratory Data Analysis*; Duxbury Press: Boston, 1981.
(15) Sebastian, P. R.; Booth, D. E.; Hu, M. Y. Using Polynomial Smoothing and Data Bounding For the Detection of Adverse Process Changes in a Chemical Process. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 881−89.
(16) Prasad, S.; Sooth, D. E.; Hu, M. Y.; Deligonul, S. The Detection of Nuclear Material Losses. *Decis. Sci.* **1995**, *26* (2), 265−81.
(17) Sebastian, P. R. Polynomial Smoothing and Data Bounding Applied to Industrial Process Control and Nuclear Material Safeguards. Ph.D. Dissertation, Kent State University, Kent, OH, 1994.

CI960044K