Research Report
Statistical Research Unit
Göteborg University
Sweden

Graphical evaluation of statistical surveillance

Marianne Frisén
Mattis Gottlow

Research Report 2003:10
ISSN 0349-8034
Revised 2005

Mailing address:	Fax	Phone	Home Page:
Statistical Research Unit	Nat: 031-773 12 74	Nat: 031-773 10 00	http://www.stat.gu.se/stat
P.O. Box 660	Int: +46 31 773 12 74	Int: +46 31 773 10 00	
SE 405 30 Göteborg			
Sweden			

Graphical Evaluation of Statistical Surveillance

Marianne Frisén and Mattis Gottlow

Statistical Research Unit, Göteborg University, Sweden

Summary. A computer program which simultaneously gives graphical information on important characteristics of statistical surveillance methods is presented. Surveillance, that is continual observation of a time series with the goal of timely detection of possible important changes in the underlying process, is used in quality control, economics, medicine and other fields. When surveillance is used in practice it is necessary to evaluate the method in order to know which action is appropriate at an alarm. The probability of a false alarm, the probability of successful detection and the predictive value are three measures (besides the usual ARL), which are illustrated by the program.

Keywords: Graphical evaluation; Computer program; Quality control; Control charts; Predicted value; Performance.

1 Introduction

Statistical surveillance is used when we have a continual observation of a time series, with the goal of detecting an important change in the underlying process as soon as possible after it has occurred and at the same time keep the rate of false alarms at an acceptable level. Statistical surveillance is used in many different fields, e.g. industrial quality control, economics, medicine and environmental control. Some examples are: Public health surveillance (Sonesson and Bock (2003),Andersson (2003), Frisén and Sonesson (2005) and Sonesson and Frisén (2005)), monitoring of a foetal heart rate during labour (Frisén (1992)) , post marketing surveillance of adverse drug effects (Lao et al. (1998)), the detection of a turn in a business cycle (Andersson et al. (2005)) and financial decision strategies (Bock et al. (2004).

Some other names of statistical surveillance are statistical process control, monitoring and change point detection. For surveys of methods, se e. g. Zacks (1983), Wetherill and Brown (1991), Basseville and Nikiforov (1993), Lai (1995) or Frisén (2003).

If a process is monitored, the decision whether a change in the process has occurred or not has to be made sequentially. Thus, the usual measures of a test's performance, the significance level and power, have to be generalized to take into account the dependence on the length of the period of surveillance and the time point where the change occurs.

A measure that is often used in quality control is the average run length (ARL) until an alarm occurs. However, ARL-curves do not contain all information about the methods. Several authors, e.g., Zacks (1980), Woodall (1983), Crowder (1987, Yashchin (1989) have pointed out that only one summarizing measure is not enough.

Address for correspondence: Marianne Frisén, Statistical Research Unit, Göteborg University, Box 660, SE 405 30 Göteborg, Sweden.
E-mail: Marianne.Frisen@statistics.gu.se

The program described in this report uses three measures of performance suggested by Frisén (1992), besides the ARL. These will give information on the influence of time and the different risks of false judgements involved when repeated decisions are made.

A description of the methods and evaluation measures used in the program are given in section 2 and 3 respectively. A description of the program, how to use it and some examples of evaluations are given in section 4.

For updates see the HTML help in the program.

2 The process under surveillance

The observations X(t) under surveillance may be averages, recursive residuals, measures of variation or some other derived statistics at time t. The variables X(t) are usually assumed independently normally distributed with constant variance. It is usually assumed that if a change in the process occurs, the mean suddenly shifts to another constant level and remains on this new level. The description below is for that case. See however the description of the SRnp method where the case of a gradual change is described.

A random process: $\{X(t): t=1,2,...\}$ is monitored and an unobserved random process $\{\mu(t): t=1,2,...\}$ determines the state of the monitored process. As mentioned above, the observations under surveillance may be some derived statistics at time t.

At time τ the monitored process changes, that is $\mu(\tau-1) \neq \mu(\tau)$. In the simplest case of two possible states we have $\mu(t)=\mu^0$ for $t=1,\ldots,\tau-1$ and $\mu(t) = \mu^1$ for $t = \tau,\ldots$ (assuming that the monitored process stays in state μ^1 once it has changed). The distribution of the random variable τ is often assumed to be known with, e.g. $\pi_t = P(\tau=t) = p(1-p)^{t-1}$. The intensity of a change is $\nu_t = P(\tau=t|\tau \geq t)$ and this probability is assumed constant over time. A common case (and the case considered in the computer program) is that the observed process is assumed to be an independent normally distributed process, whose mean value changes from μ^0 to μ^1. The variance of the process remains unchanged:

$$X(t) \sim \begin{cases} N(\mu^0,\sigma^2) & \tau \leq t \\ N(\mu^1,\sigma^2) & \tau > t \end{cases}$$

The size of the shift, $|\mu^1 - \mu^0|$, as well as the intensity of a change, is affecting the performance a surveillance method. Several interesting sizes of both the shift and the intensity can be tried in the program. The variance is by scaling set to one.

For the SRnp method the shift is a change in slope as described in the section for that method.

3 Methods

Two specific methods of surveillance often used in quality control are now available in the computer program: the Shewhart and the CUSUM methods. Also available is a method based on the likelihood ratio, the Shiryaev-Roberts method. These are all parametric methods, but the program also presents a new robust method, developed at the department, the SRnp (Shiryaev-Roberts non-parametric) method that is used to detect turning points in cyclical processes.

For some methods in the program there is a choice between one- and two-sided methods. A twosided method is used when a process is monitored to detect either a positive (upward) or a negative (downward) shift. The methods described below are all described for the one-sided case, but the description can easily be adapted for the two-sided case.

3.1 Shewhart

The Shewhart method was suggested already in 1931 and it is described in all elementary books on quality control. At decision time t only the last observation X(t) is considered. This observation could be a standardized mean or another derived statistic. This method can be regarded as repeated ordinary tests of hypotheses based on the information at the last time point. An alarm is triggered at time t if the distance between X and the target value exceed a limit *L*:

$$t_A = \min\{s; X(s) > L\}$$

A standard value is *L*=3.

3.2 CUSUM

Page (1954) suggested the CUSUM method and it is described in most books on quality control. The cumulative sums C(*i*) of differences between the observations and the target value are calculated for *i* = 1,...,t and C(0) is set to zero. There is an alarm at time t if C(t)-C(t-*i*) - *ki* is greater than the limit *L* for any *i*:

$$t_A = \min\{s; C(s) - C(s-i) - ki > L \text{ for some } i = 1,\dots,s\}$$

The properties of this method are determined by the value of the parameters *k* and *L*. The information from earlier observations is handled differently depending on the position in the time series. Recent observations have more weight than old ones.

The test might be performed by moving a V-shaped mask over a diagram until any earlier observation is outside the limits of the mask. Thus the method is often referred to as "the V-mask method". The parameter *L* determines the distance between the last observation and the apex of the "V". The parameter *k* determines the slopes of the legs. It might be of interest to try several pairs of parameters *L* and *k*. However, in this version of the program only certain values can be chosen.

3.3 SR

The Shiryaev-Roberts method, SR, is a variant of the full likelihood method, LR.

The LR method was constructed by Frisén and de Maré (1991) to meet several optimality criteria. The general method uses combinations of partial likelihood ratios:

$$L(s,t) = \ln \frac{f_{X_t|\tau=s}(x_t)}{f_{X_t|\tau>s}(x_t)}$$

where $x_t=\{x(i); i\leq t\}$. If the process under surveillance is a sequence of iid variables (within each state) we get:

$$L(s,t) = \sum\nolimits_{i=s}^{t} \ln \frac{f_1(x(i))}{f_0(x(i))}.$$

The Shiryaev-Roberts method can be regarded as a version with a non-informativ prior for the change point time. It can alternatively be regarded as a version with a very small probability for a change at each time point.

3.4 SRnp

The SRnp method was developed at the Department of Statistics, Gothenburg University, Sweden. First suggested by Frisén (2000) and described and evaluated by Andersson (2002). The method is developed for cyclical processes and the aim is to detect a turn (peak or trough) as soon as possible. The method is based on the likelihood ratio.

Since the characteristics of a cyclical process (cycle length and amplitude) changes over time it is very difficult to find a suitable parametric model for the cycles. Frisén (2000) suggested that a completely non-parametric estimation procedure could be used, not based on any parametric model but only on monotonicity restrictions. The estimation procedure is the regression under order-restrictions. The regression function is estimated under the restriction of monotonicity (no turn), see Robertson et al. (1988), and under the restriction of unimodality (turn), see Frisén (1986):

No turn: $\mu(1) \leq \ldots \leq \mu(s)$

Turn: $\mu(1) \leq \ldots \leq \mu(\tau\text{-}1)$ and $\mu(\tau\text{-}1) \geq \mu(\tau) \geq \ldots \geq \mu(s)$

The advantage of this approach is that miss-specifications of the cycles are avoided. It was shown in Andersson et al. (2005) that even a small miss-specification of the slope before the turn leads to bad properties of the surveillance system (long delay time, low predictive value).

An alarm is triggered as soon as the likelihood ratio exceeds a limit *L*. This means that similarity of the observations to a unimodal pattern compared to a monotonic pattern is compared.

Since the SRnp methods is developed for cyclical processes the process parameter “Shift” in the program does not denote the size of a shift from one constant level to another, as is the case for the other methods. Instead, it denotes the change in the average slope. The value is standardized for a process with variance 1.

4 Evaluation

The proper evaluation of surveillance methods enables us to choose the method which is most suited for the problem at hand. Optimality of methods for surveillance is discussed in e.g. Frisén and de Maré (1991) and Frisén (2003).

In order to fully evaluate a surveillance method several measures are useful, some of which are presented here. These measures also give information on which action is appropriate when we have an alarm.

4.1 ARL

Average Run Length: A measure which is often used in quality control is the average run length (ARL) until an alarm occurs. ARL_0 is the average number of runs before an alarm when there is no change in the system under surveillance. The average run length under the alternative hypothesis, ARL_1, is the mean number of decisions that must be taken to detect a change that occurred at the same time as the inspection started.

Roberts (1966) and Goel and Wu (1971) have given very useful diagrams of the ARL. However, it is recommended that also other measures of the performance are used. Sometimes the median is used instead of the average, so that we have MRL_0 as the median number of runs before an alarm is called, when there is no change in the system and MRL_1 is the mean number of decisions that must be taken to detect a change that occurred at the same time as the inspection started.

For a Shewhart test, exact calculation is simple; ARL = $1/p$, where p = $P(|X(t)| > L)$ for the two-sided method and p = $P(X(t) > L)$ for the one-sided method. For the other methods, large simulations were used to determine the ARL.

ARL_0 (as well as MRL_0) depends only on the method and its parameters. ARL_1 (as well as MRL_1) depends also on the size of the shift.

4.2 Alarm probabilities

The probability of false alarm (PFA) is defined as the probability of an alarm no later than at time t, given that no change has occurred:

$$PFA(t) = P(t_A \leq t | \tau > t)$$

In cases where the surveillance is not stopped until there is an alarm, the total probability of a false alarm is equal to one. This is the case for the methods used here and for most other methods. Thus the curve has an asymptote at one. In some cases the length of the surveillance time is limited and the curve can be used to give the total false alarm probability for different lengths.

The method and its parameters determine the probability of false alarm.

4.3 Probability of successful detection

The probability of successful detection (PSD) is the probability to get an alarm within d time units after the change has occurred, conditioned that there was no alarm before the change:

$$PSD(\tau, d) = P(t_A - \tau < d \mid t_A \leq \tau)$$

To calculate the Probability of Successful Detection it is necessary to specify the time interval, named d, within which detection is desired. Often the time between the change and the detection is crucial for the possibility of rescuing action. Several values of d can be tried in the program.

4.4 Predictive Value

The predictive value (PV) of an alarm at time t is the probability that a change has occurred at t or later, given that there was an alarm at t:

$$PV(t) = P(\tau \leq t | t_A = t)$$

It is thus the relative frequency of motivated alarms among all alarms at a certain point of time. It gives information on whether an alarm is a strong indication of a change or not.

The formula for the predictive value and its asymptote are found in Frisén (1992) for the Shewhart method.

To calculate the predictive value not only the method and its parameters but also the intensity and the size of the shift have to be specified.

5 The program

The program is written using Visual Basic 6.0. An earlier program with similar content was made by Claes Cassel in cooperation with Marianne Frisén. Improvements of the program was made in 2005 by Linus Schiöler.

By setting different values of the parameters, the user can visually explore the effects of different conditions for different methods. In the present version the Shewhart, CUSUM, SR and the SRnp methods are available. Explanations and help texts are given in HTML documents, which are available from the Help menu.

Three functions are shown in three graphs. These are the probability of false alarm, the probability of successful detection and the predictive value. Different settings of parameters results in different functions and the results can be compared in the same graphs.

The evaluation functions are computed by the methods in Frisén (1992). For the Shewhart method the calculations are exact but for the other methods they are computed externally by large scale simulations.

The program runs under Windows and about 10 Mb of disk space is needed to install the program.

Installation instructions:

1. Close all applications.
2. Run the setup.exe file from the CD
3. Follow the instructions on the screen.
4. If you have an earlier version installed you will get help to remove that. Then start the installation again.
5. Run the program from Start>Programs>SE 1.3 or from the shortcut on the desktop.

5.1 Using the program

To create graphs, one has to provide the following input (preferably in that order):

1. *Choose Method.* The following methods are currently available:
   - Shewhart
   - CUSUM
   - Shiryaev-Roberts
   - SRnp

2. *Chose method parameters.*
   - Shewhart: The Limit *L* and *Two-sided /One-sided.*
   - CUSUM: The Limit *L*, the *k1* parameter and *Two-sided /One-sided.*
   - SR: The Limit *L*, (*Two-sided* is not available)
   - SRnp: The Limit *L*, (*Two-sided* is not available)

3. *Specify the problem*
   - Size of the Shift. It is assumed that if a change in the process occurs, the mean suddenly moves to another constant level and remains on this new level. The size of the shift, that is absolute value of the difference between the target level and the new level, is denoted *m* in the graph legends. This value is needed for the calculation of both PSD and PV.
   - Intensity. To calculate the PV the incidence of changes has to be specified. The incidence of a change (sometimes named the intensity) is here abbreviated as *i*. It is the probability that the change occurs at a certain time, given that it has not occurred before. The incidence is assumed constant over time
   - Detect within *d*. To calculate the PSD, it is necessary to specify a time interval, named *d*, within which detection is desired.

4. *Choose the number of time points.* For each of the three types of graphs you can choose the number of time points that you want to have displayed in the plot. When you have more then one curve per plot, all curves must have the same number of time points.

5. *Create the graphs.*
   - Add to graphs. If the *Add to Graph* button is pressed, curves corresponding to the chosen input are created in the graphs. If a plot has been created previously, the new curve is added to the existing plot. The chosen parameter values and the corresponding ARL_0 and ARL_1 are also displayed.
   - New Graphs. If the *New Graph* button is pressed, all output is deleted before the results are displayed

- Graph legends. The parameter values *that affect the different kinds of graphs* are given in the graph legends. An '*' indicates that the method is *twosided.*
- The *Method list*: In this list, all the chosen parameter values for the created plots are displayed. An '*' indicates that the method is *twosided.*

Some notes:
For methods other than the Shewhart method, data for the graphs have to be calculated using Monte Carlo simulations. This means that if a non-Shewhart method is chosen, data are read from certain tables and one has only a limited set of parameters to choose from.
The variance of the process under surveillance is by scaling set to one.
The *decimal delimiter used is a comma sign.* If a decimal point is used it is automatically changed to a comma sign.

Saving, loading and printing

- From the *File* menu the data from the three graphs can be saved as a .gda file, by selecting *Save Graph Data.* This file can only be opened by the SE program (by selecting *Load Graph Data*). The three graphs can only be saved or loaded all at the same time. When a file is opened, additional methods can be included in the graphs as described above.
- Printing: Double-clicking on a graph makes it appear in a new window, from where it can be printed by clicking on the *Print* menu.
- When a graph is opened in a new window, the graph data can be copied to the clipboard by clicking on the *Copy Data* menu. The data can then be pasted into other applications (e. g. Excel) for further analysis.

5.2 Examples of evaluations

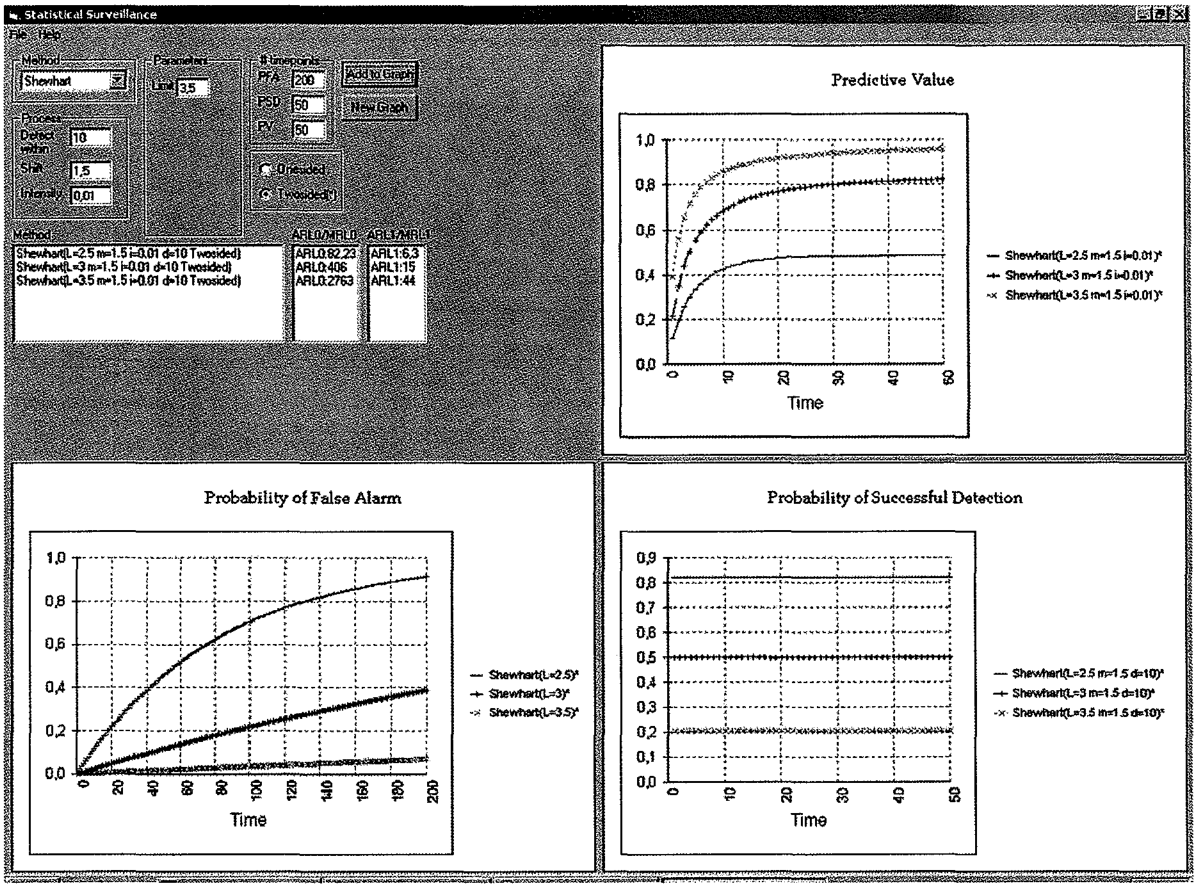

Figure 1. Evaluations of the Shewhart method for different values of *L*. Observe that the harder it is to get an alarm the better is the predicted value of an alarm.

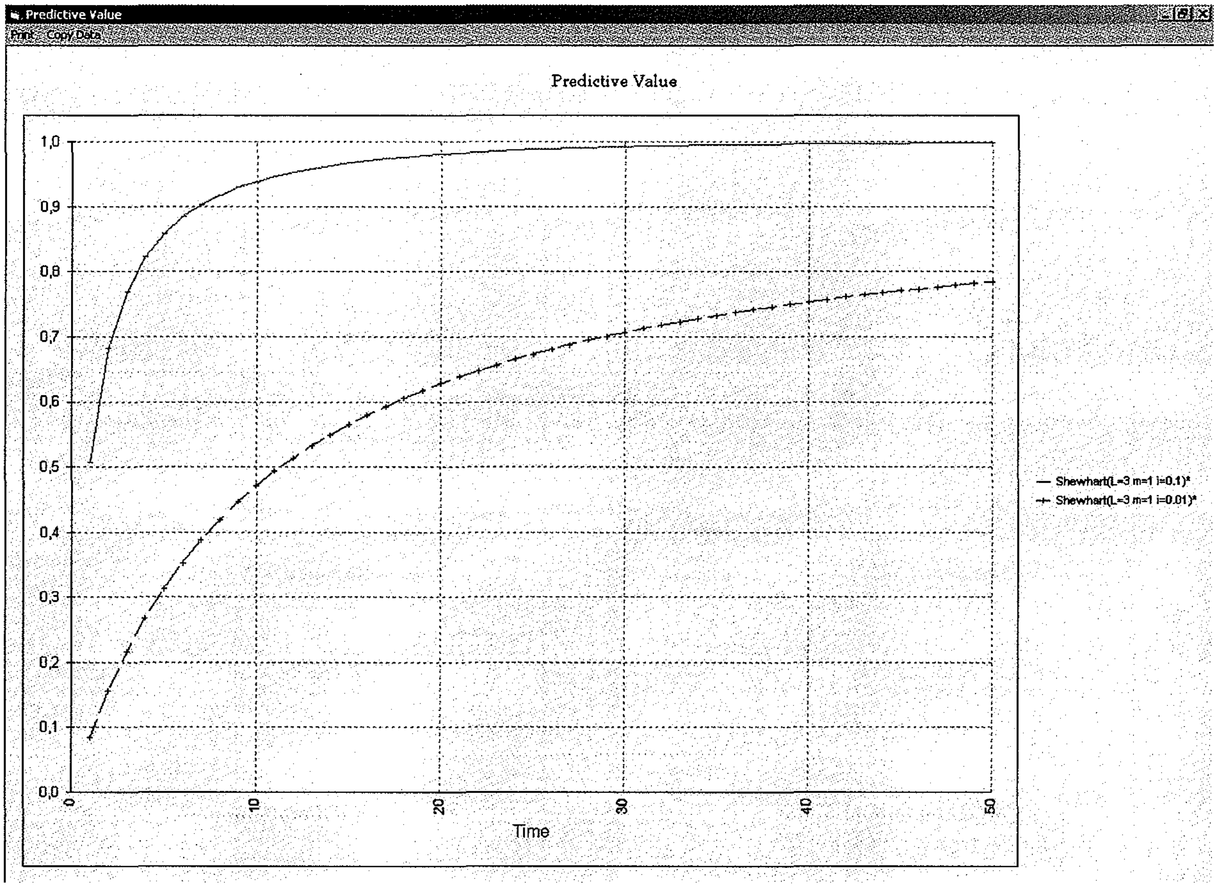

Figure 2. Detailed picture of the predictive value of the Shewhart method for different intensities. *Shift* = 1, *L* = 3. The predicted value of an alarm is higher if the intensity of changes is higher.

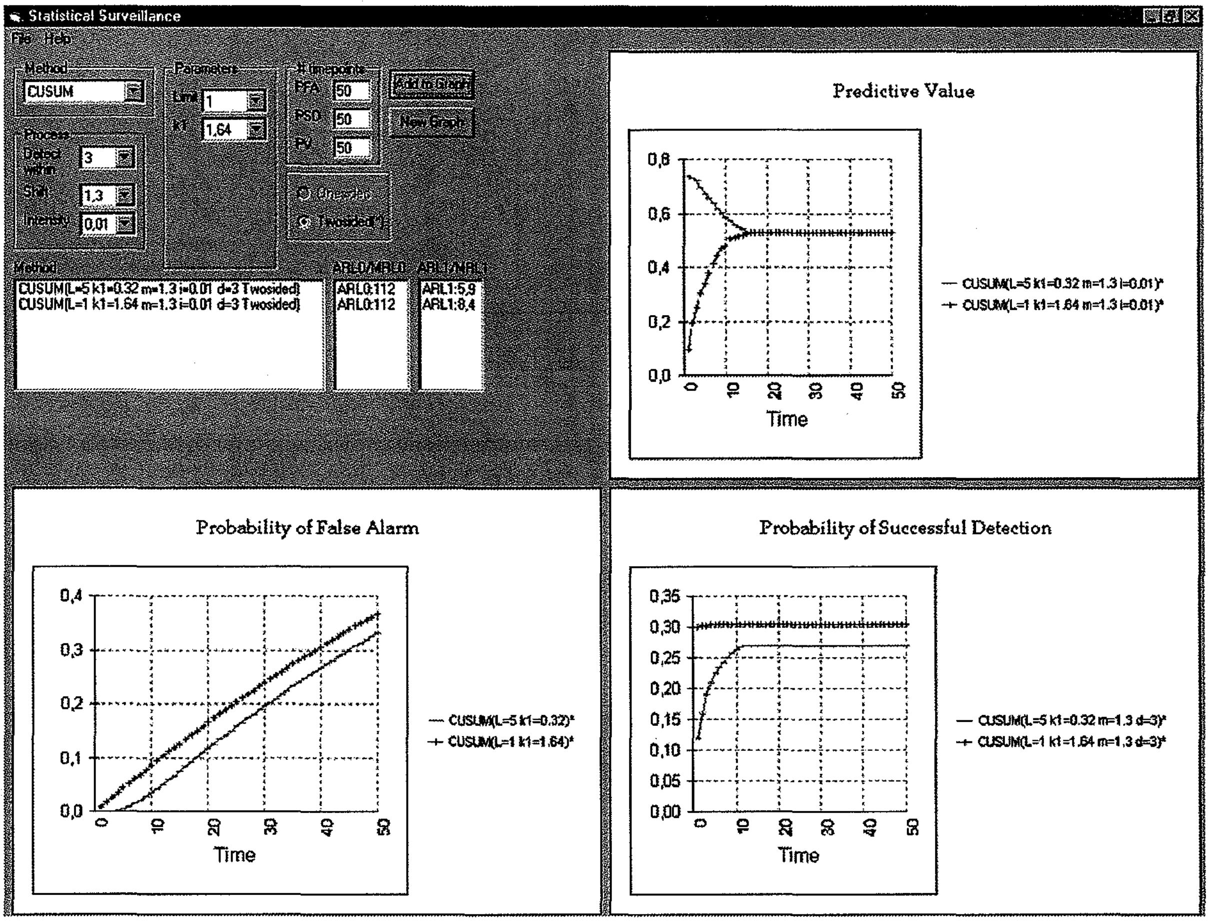

Figure 3. Two variants of the CUSUM method are compared. They have different combinations of the parameters L and k but in such a way that the ARL_0 is the same.

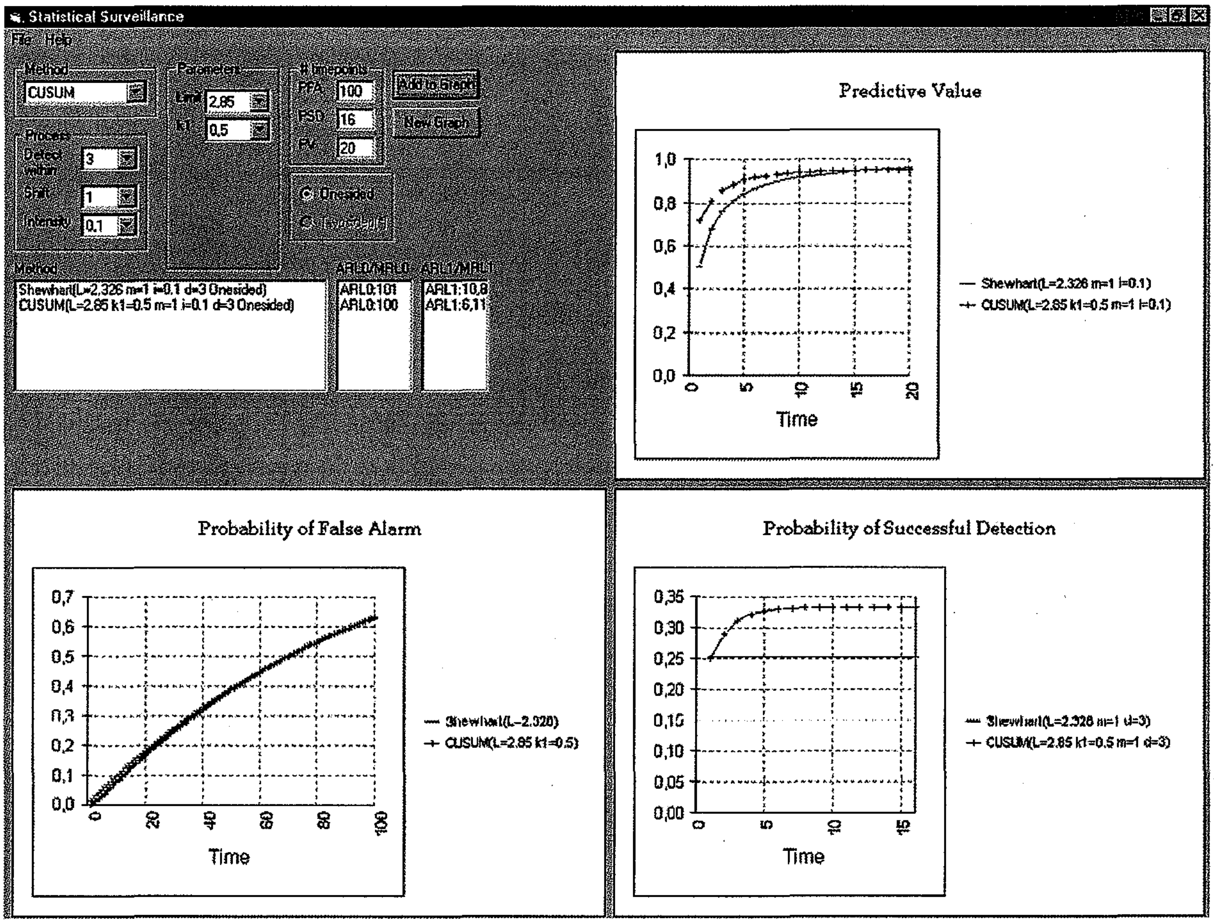

Figure 4. An example of evaluation of the Shewhart and the CUSUM method. The methods have here the same ARL_0 but the properties of the methods differ substantially. Observe that for the Shewhart method the predicted value is low for early alarms but better for later alarms.

6 Conclusion

By trying different values of the parameters you get help to choose the right parameter for your application. By varying one variable at a time you will also get insight in how this variable influences the different characteristics.

References

Andersson, E. (2002) Monitoring cyclical processes - a nonparametric approach. *Journal of Applied Statistics,* **29,** 973-990.

Andersson, E. (2003) A monitoring system for detecting starts and declines of influenza epidemics. *Morbidity and Mortality Weekly Report.*

Andersson, E., Bock, D. and Frisén, M. (2005) Statistical Surveillance of Cyclical Processes with Application to Turns in Business Cycles. *Journal of Forecasting.*

Basseville, M. and Nikiforov, I. (1993) *Detection of Abrupt changes- Theory and Application,* Prentice Hall.

Bock, D., Andersson, E. and Frisén, M. (2004) The relation between statistical surveillance and certain decision rules in finance. *Submitted.*

Crowder, S. (1987) A simple method for studying run-length distributions of exponentially weighted moving average charts. *Technometrics,* **29,** 401-407.

Frisén, M. (1986) Unimodal regression. *The Statistician,* **35,** 479-485.

Frisén, M. (1992) Evaluations of Methods for Statistical Surveillance. *Statistics in Medicine,* **11,** 1489-1502.

Frisén, M. (2000) Statistical Surveillance of Business Cycles. 1994:3 Revised, Department of Statistics, Göteborg University.

Frisén, M. (2003) Statistical surveillance. Optimality and methods. *International Statistical Review,* **71,** 403-434.

Frisén, M. and de Maré, J. (1991) Optimal Surveillance. *Biometrika,* **78,** 271-80.

Frisén, M. and Sonesson, C. (2005) Optimal surveillance. In *Spatial surveillance for public health* (Eds, Lawson, A. and Kleinman, K.) Wiley, New York.

Goel, A. L. and Wu, S. M. (1971) Determination of ARL and a Contour Nomogram for Cusum Charts to Control Normal Mean. *Technometrics,* **13,** 221-230.

Lai, T. L. (1995) Sequential Changepoint Detection in Quality-Control and Dynamical-Systems. *Journal of the Royal Statistical Society B,* **57,** 613-658.

Lao, C. S., Kessler, L. G. and Gross, T. P. (1998) Proposed statistical methods for signal detection of adverse medical device events. *Drug Information Journal,* **32,** 183-191.

Page, E. S. (1954) Continuous inspection schemes. *Biometrika,* **41,** 100-114.

Roberts, S. W. (1966) A Comparison of some Control Chart Procedures. *Technometrics,* **8,** 411-430.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988) *Order Restricted Statistical Inference,* John Wiley & Sons Ltd.

Sonesson, C. and Bock, D. (2003) A review and discussion of prospective statistical surveillance in public health. *Journal of the Royal Statistical Society A,* **166,** 5-21.

Sonesson, C. and Frisén, M. (2005) Multivariate surveillance. In *Spatial surveillance for public health* (Eds, Lawson, A. and Kleinman, K.) Wiley, New York, pp. 169-186.

Wetherill, G. B. and Brown, D. W. (1991) *Statistical process control,* Chapman and Hall.

Woodall, W. H. (1983) The Distribution of the Run Lenght of One-Sided CUSUM Procedures for Continous Random Variables. *Technometrics,* **25,** 295-301.

Yashchin, E. (1989) Weighted Cumulative Sum Technique. *Technometrics,* **31,** 321-338.

Zacks, S. (1980) Numerical determination of the distributions of stopping variables associated with sequential procedures for detecting epochs of shift in distributions of discrete random variables. *Communications in Statistics. Sim & Comput,* **9,** 1-18.

Zacks, S. (1983) Survey of classical and Bayesian approaches to the change-point problem: Fixed sample and sequential procedures of testing and estimation. In *Recent advances in statistics*, pp. 245-269.