



UNIVERSITY OF GÖTEBORG

Department of Statistics

RESEARCH REPORT 1994:3
ISSN 0349-8034

VISUAL EVALUATIONS OF STATISTICAL SURVEILLANCE

by

Marianne Frisé
&
Claes Cassel

Statistiska institutionen
Göteborgs Universitet
Viktoriagatan 13
S-411 25 Göteborg
Sweden

VISUAL EVALUATIONS OF STATISTICAL SURVEILLANCE

M Frisé¹ and C Cassel²

A computer program which simultaneously gives visual information on important characteristics 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 self-instructive computer program.

KEY WORDS: Visual presentation; Computer program; Quality control; Control charts; Predicted value; Performance;

¹Department of Statistics, Göteborg University, Viktoriagatan 13, S-411 25 Göteborg, Sweden.

²Department of Economic Statistics, Stockholm School of Economics, Box 6501, S-11383 Stockholm, Sweden.

1. INTRODUCTION

Continual surveillance to detect some event of interest, is used in many different areas, e.g. industrial quality control, detection of shifts in economic time series, medical intensive care and environmental control. For applications in medicine, see Friséen (1992) and for applications in economics, see Friséen (1994a).

For surveys of methods, see e.g. Zacks (1983), Wetherill and Brown (1990) and Friséen (1994b). Some methods (like the Shewhart test) only take the last observation into account. Others (simple sums or averages) give the same weight to all observations. For most applications it is relevant to use something in between. That is, all observations are taken into account but more weight is put on recent observations than on old ones. The CUSUM method is an example of such a method. The relative weight on recent observations and old ones can be continuously varied by varying its two parameters. Both the Shewhart and the CUSUM method have certain optimality properties as described by Friséen and de Maré (1991) and Friséen (1994c).

The usual measures of a test's performance, namely the significance level and the power, would have to be generalized in any of many possible ways to take into account the dependence on the length of the period of surveillance and the time point t' where the change occurs. Here, a visual demonstration which shows how these variables (which vary between practical situations) influence different measures (for different methods) is presented.

A measure which is often used in quality control is the average run length (ARL) until an alarm occurs. It was suggested already by Page (1954). Roberts (1966) has given very useful diagrams of the ARL. Later several authors e.g. Saccucci and Lucas (1990), Champ and Rigdon (1991), Champ et.al. (1991), Yashchin (1992) and Yashchin (1993) have studied the ARL of specific methods and models. However, ARL-curves do not contain all information about the methods. Several authors e.g. Zacks (1980), Woodall (1983), Crowder (1987) and Yashchin (1989) have pointed out that only one summarizing measure is not enough.

This program uses three measures of performance suggested by Friséen (1992) besides the ARL. These will give information on the influence of time and the different risks of false judgements involved when repeated decisions will be made.

The program description and operation are given in Sections 2 and 3. Help texts are exemplified in Section 4. The situations considered and the necessary input are described. There are also texts on some methods of surveillance and the measures to be used in the evaluations. The output for some examples are given and discussed in Section 5.

2. PROGRAM DESCRIPTION

The program is written using the Realizer programming language, which is a dialect of Basic compatible to Q-basic. Realizer is similar to Microsoft's Visual Basic. It runs under Windows. The graphical user interface objects in Windows such as for example buttons, scrollbars, menus and charts are used to build screen interfaces designed for easy interaction between the user and the program.

By setting different values of the parameters the user can explore visually the effects of different conditions for different methods. In the present version the Shewhart- and the CUSUM -methods are available. Explanations and help texts are displayed in screens which are available by clicking a button.

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 the parameters result in different functions. If desired, the results can be compared in the same graphs and printed.

The program Statistical Surveillance runs under Windows on PC:s. A mouse and 0.5 Mb hard disk space is needed.

This program consists of four parts:

1. A control program which controls the main flow and what is shown in the different screens.
2. The main program for calculations which serves as a basis for the evaluations.
3. A program containing the different screens.
4. A program containing the texts to be displayed in the screens.

The evaluation functions are computed by the methods given in Frisé (1992). For the Shewhart method and the first time points of the CUSUM method the calculations are exact. Other values are computed externally by large scale simulations.

The program is available free from the authors.

3. PROGRAM OPERATION

The program is self-instructing with help texts. Examples will be given below. The introductory text which is presented in Figure 1 gives instructions for the general operation. Details on the operation are given in the help text on controls. See Figure 2.

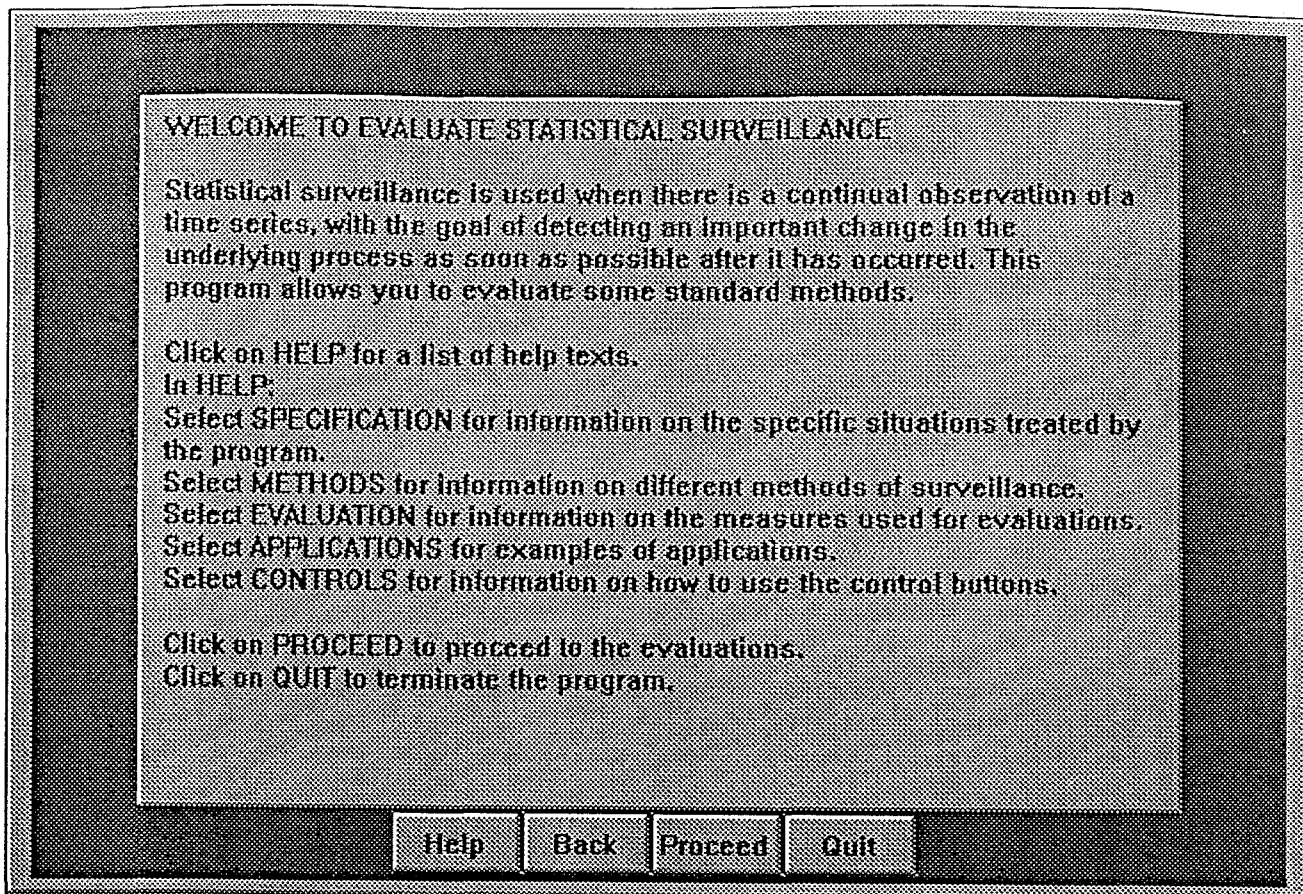


Figure 1. The introductory screen.

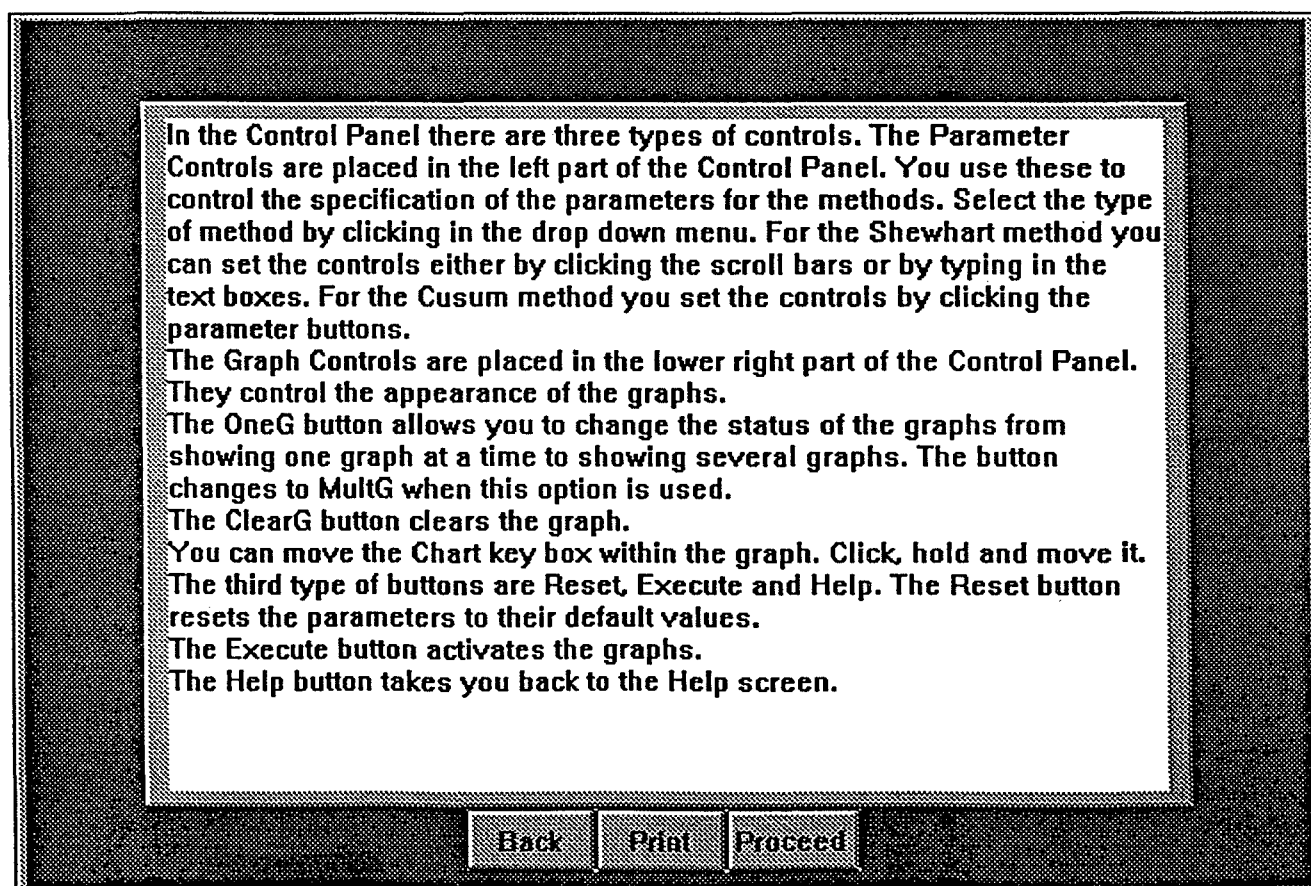


Figure 2. A click on **CONTROLS** gives these instructions for operation.

4. EXAMPLES ON HELP TEXTS

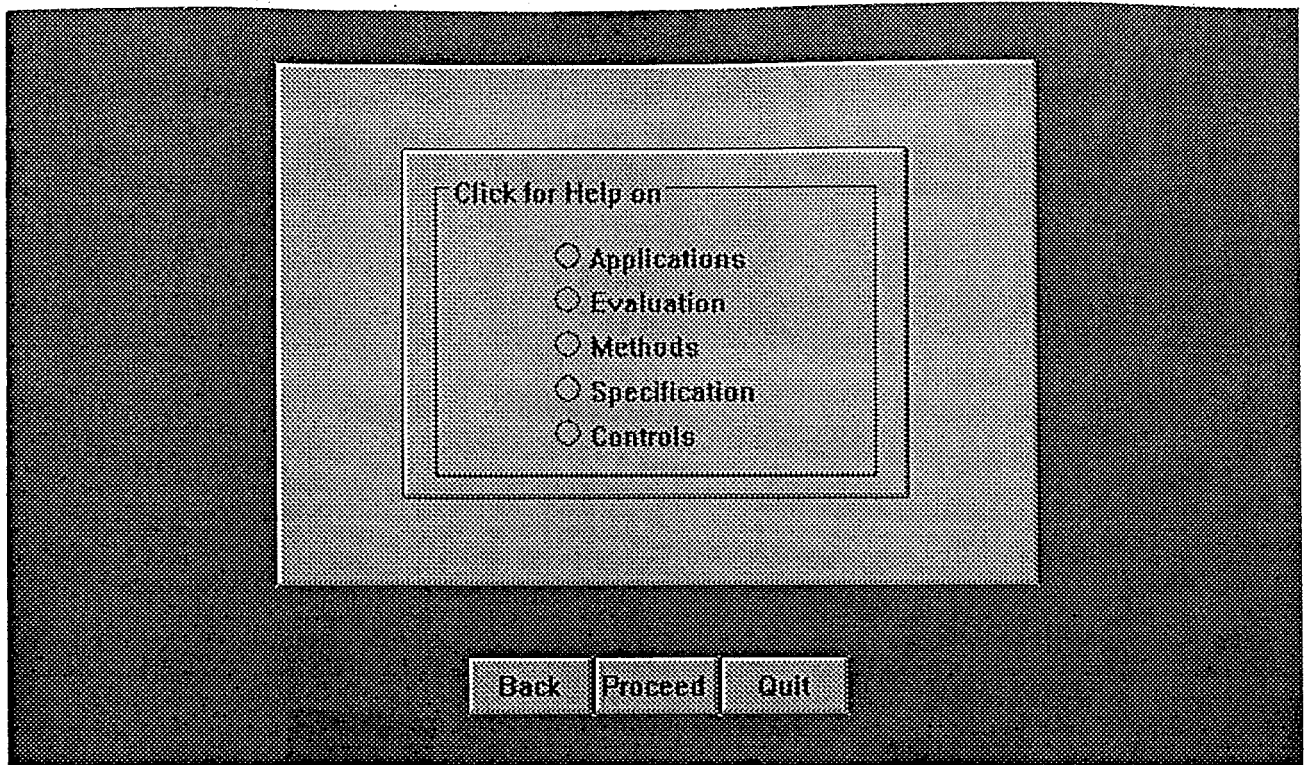


Figure 3. A click on HELP gives the first branches of the tree of help texts.

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 assumed independently normally distributed with constant variance. The variance is by scaling set to one. As is usually done 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 named "M". Two-sided methods are used. Several interesting values of M can be tried.

To calculate the probability of successful detection it is necessary to specify a 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.

To calculate the predicted value the incidence of changes has to be specified. The incidence of a change (sometimes named the intensity) is here abbreviated as "inc". 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. Several interesting values of inc can be tried.

Figure 4. A click on SPECIFICATION gives the above notations and specifications.

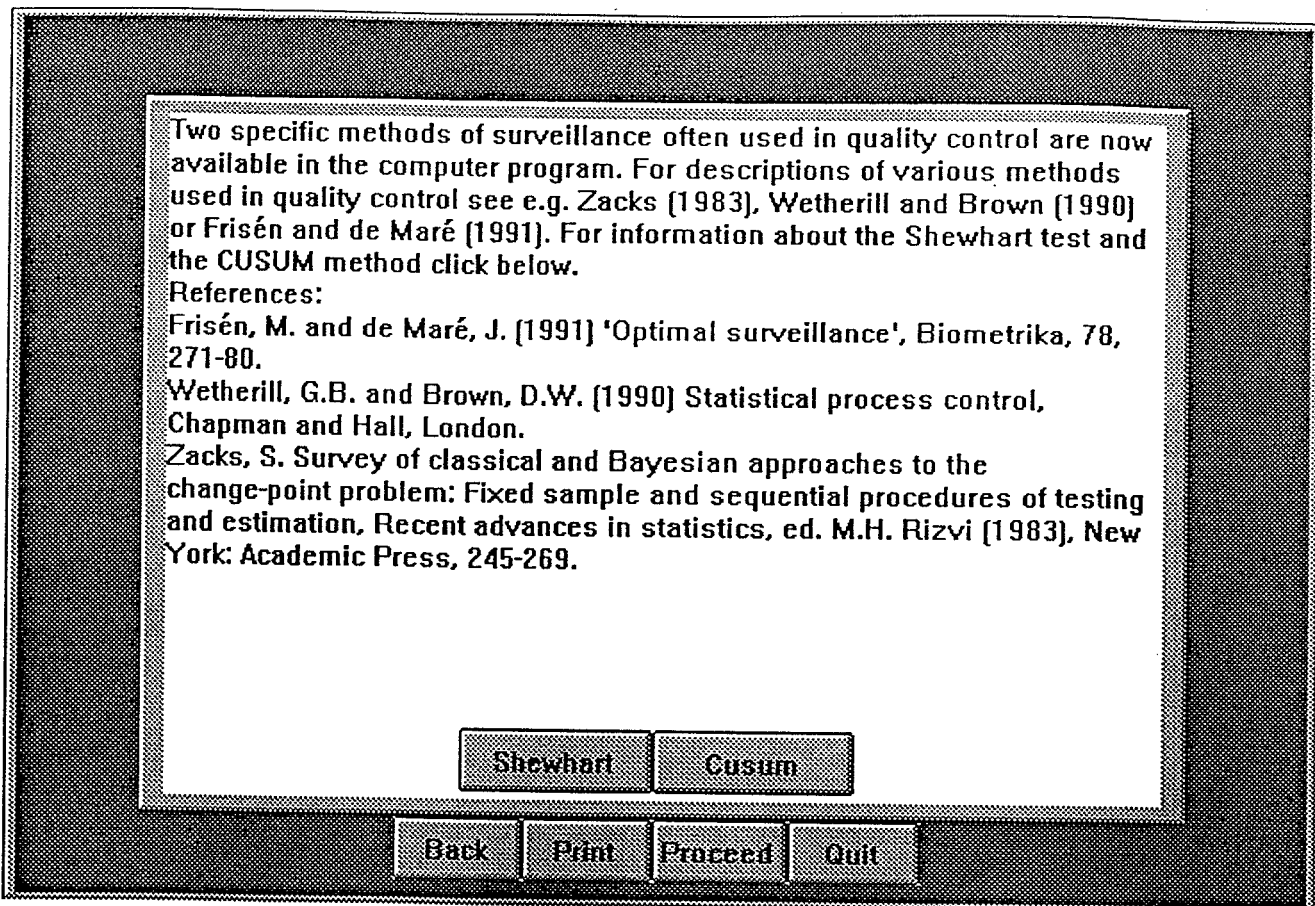


Figure 5. A click on METHODS gives the above presentation.

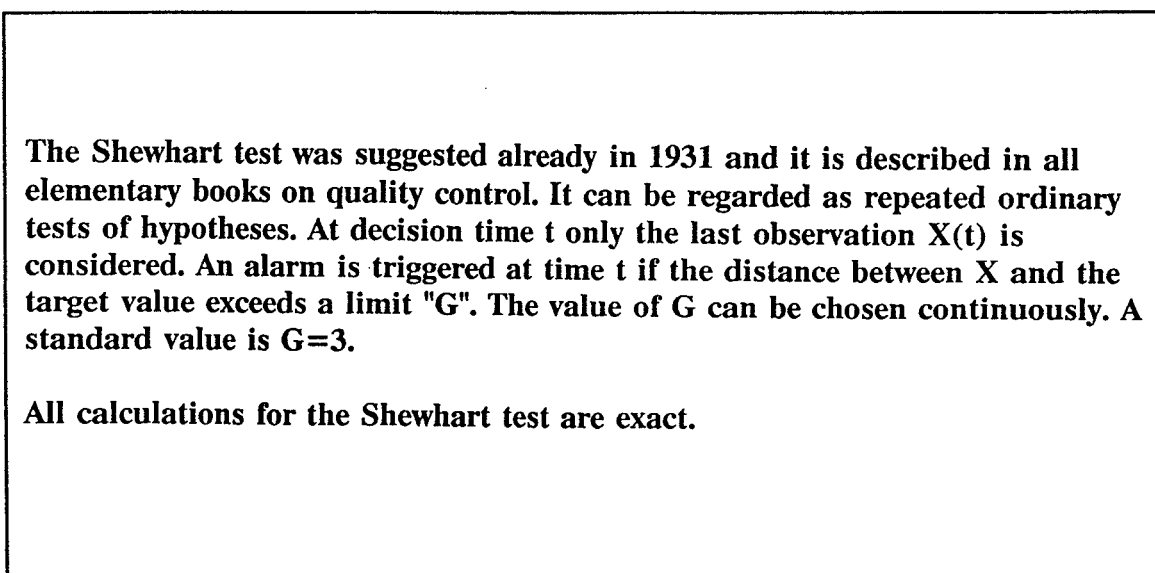


Figure 6. A click on SHEWHART gives this presentation of the method.

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 $i=1,\dots,t$ and $C(0)$ is set to zero. There is an alarm at time t if $|C(t)-C(t-i)|$ is greater than $h+ki$ for any i .

The properties of this method are determined by the value of the parameters k and h . 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 h 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 h and k . However, in this first version of the program only certain discrete values can be chosen.

The evaluations of the CUSUM method are made by large scale simulations. As many as 100 000 replicates for each point in a diagram were used in order to avoid any uncertainty because of the simulations.

Figure 7. A click on CUSUM gives the above presentation of the method.

Evaluations are necessary in order to choose the appropriate method and parameter set for a problem in practice. Also, after these choices the decision on the appropriate action after alarm should be based on knowledge on the different error rates.

The run-length distributions for all interesting cases (also those where the change appears after the start of the surveillance) contain the information necessary for an evaluation of a method or a comparison between some methods. The actual comparison is usually based on the average run length, ARL. Click the button marked ARL for more information.

Only one summarizing measure of the distributions is not enough. Because of a complicated time dependence and the dependence of the incidence other measures besides the ARL are of interest. Frisén (1992) suggested the use of the probability of false alarm, the probability of successful detection and the predictive value. Click the buttons below for information on these measures of performance.

References:
Frisén, M. (1992). Evaluations of methods for statistical surveillance, *Statistics in Medicine* 11, 1489-1502.

Buttons: ARL, Pr False Al, Pr Succ Det, Pred Val, Back, Print, Proceed, Quit

Figure 8. A click on EVALUATION gives this discussion and reference.

Continual surveillance to detect some event of interest, is used in many different areas, e.g. industrial quality control, detection of shifts in economic time series, medical intensive care and environmental control. For applications in medicine see Friséen (1992) and for applications in economics see Friséen (1994).

Friséen, M. Evaluations of methods for statistical surveillance, *Statistics in Medicine* 11, 1489-1502 (1992).

Friséen, M. Statistical surveillance of business cycles. Research report. Dept of Statistics, University of Göteborg (1994).

Figure 9. A click on APPLICATIONS gives the above references.

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.

For a Shewhart test exact calculation is simple; $ARL = 1/p$, where $p = P(|X(t)| > G)$. For the CUSUM method large simulations were used to determine the ARL.

ARL_0 depends only on the method and its parameters. ARL_1 depends also on the size M of the shift.

Goel, A.L. and Wu, S.M. 'Determination of A.R.L. and a contour nomogram for cusum charts to control normal mean', *Technometrics*, 13, 221-230 (1971).

Roberts, S.W. 'A comparison of some control chart procedures', *Technometrics*, 8, 411-430 (1966).

Figure 10. A click on ARL will give the above description and references.

The probability of false alarm is defined as the probability of an alarm no later than at time t given that no change has occurred. It corresponds to the probability of erroneous rejection of the null hypothesis, the level of significance, but is a function of the time 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 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.

Figure 11. A click on PROBABILITY OF FALSE ALARM gives the presentation above.

The distance between the change and the alarm, sometimes called the residual run length (RRL) is of interest in many cases. Several optimality conditions are based on this distance. One characterization of the distribution of the RRL is the probability that the RRL is less than a certain constant D (the time limit for successful rescuing action). This measure, the probability of successful detection, 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.

To calculate the probability of successful detection not only the method and its parameters but also the time distance, D , and the size, M , of the shift have to be specified.

Figure 12. A click on PROBABILITY OF SUCCESSFUL DETECTION gives the above presentation.

The predictive value of an alarm at time t is the probability that a change has occurred at t or before, given that there was an alarm at 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 Friséen (1992) for the Shewhart method. The fact that the curve, for certain parameters of the CUSUM method, is not smooth is not due to uncertainty in the simulations. In fact, the predicted value is not always an increasing function of the time of the alarm.

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

Frisén, M. Evaluations of methods for statistical surveillance, *Statistics in Medicine* 11, 1489-1502 (1992).

Figure 13. A click on PREDICTIVE VALUE gives the presentation above.

5. EXAMPLES ON EVALUATIONS

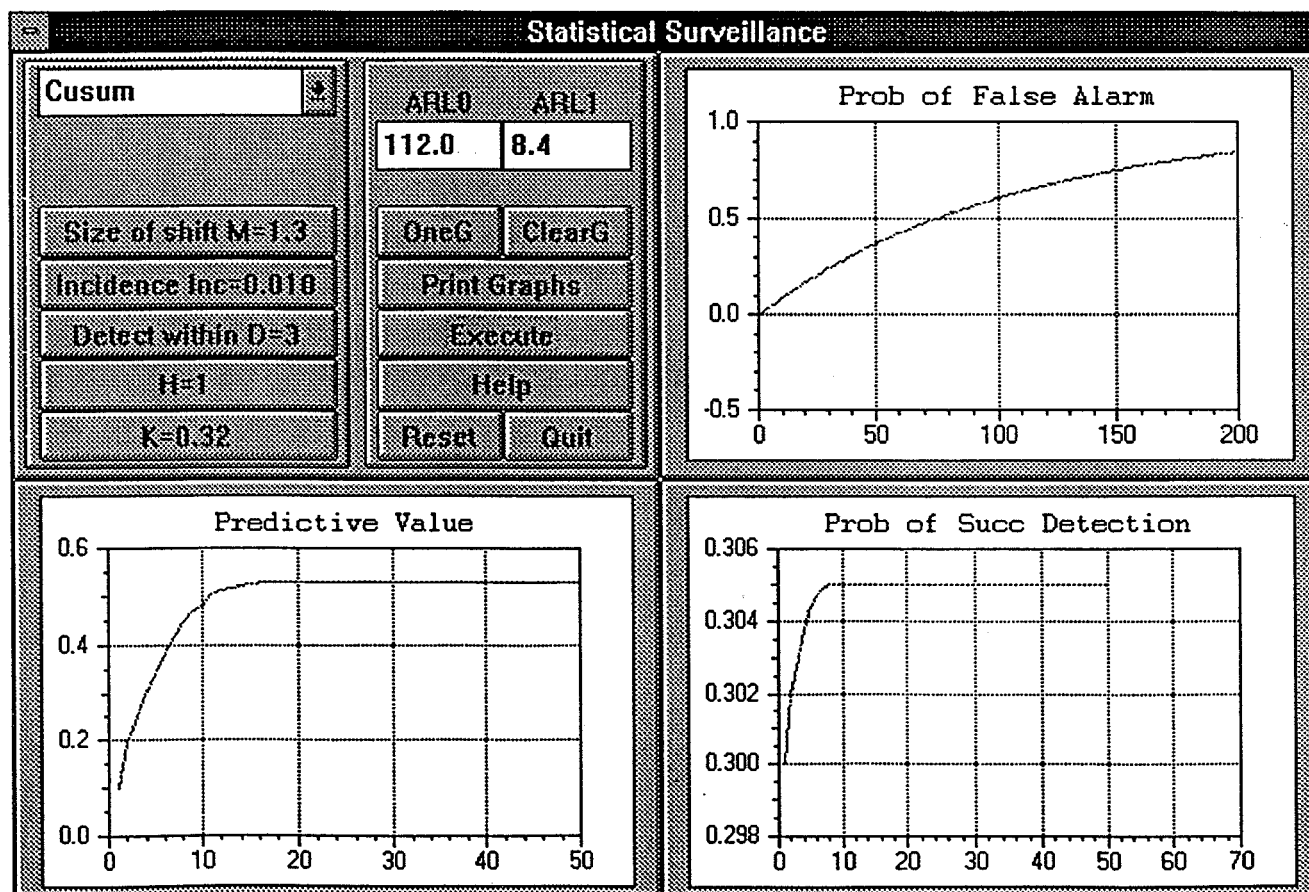


Figure 14. An example of evaluation of the CUSUM method.

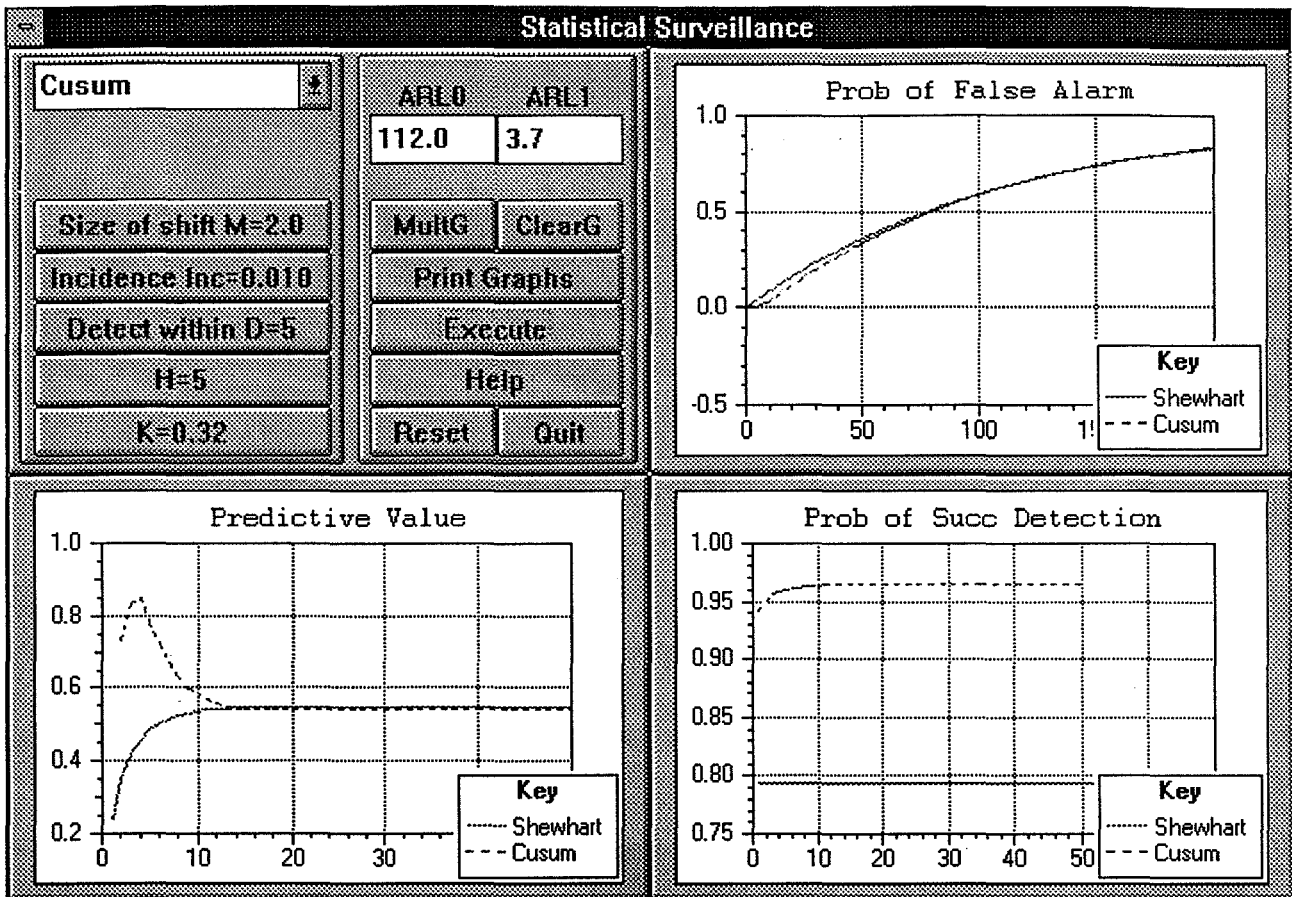


Figure 15. Evaluation of the Shewhart ($G=2.61$) and the CUSUM method.

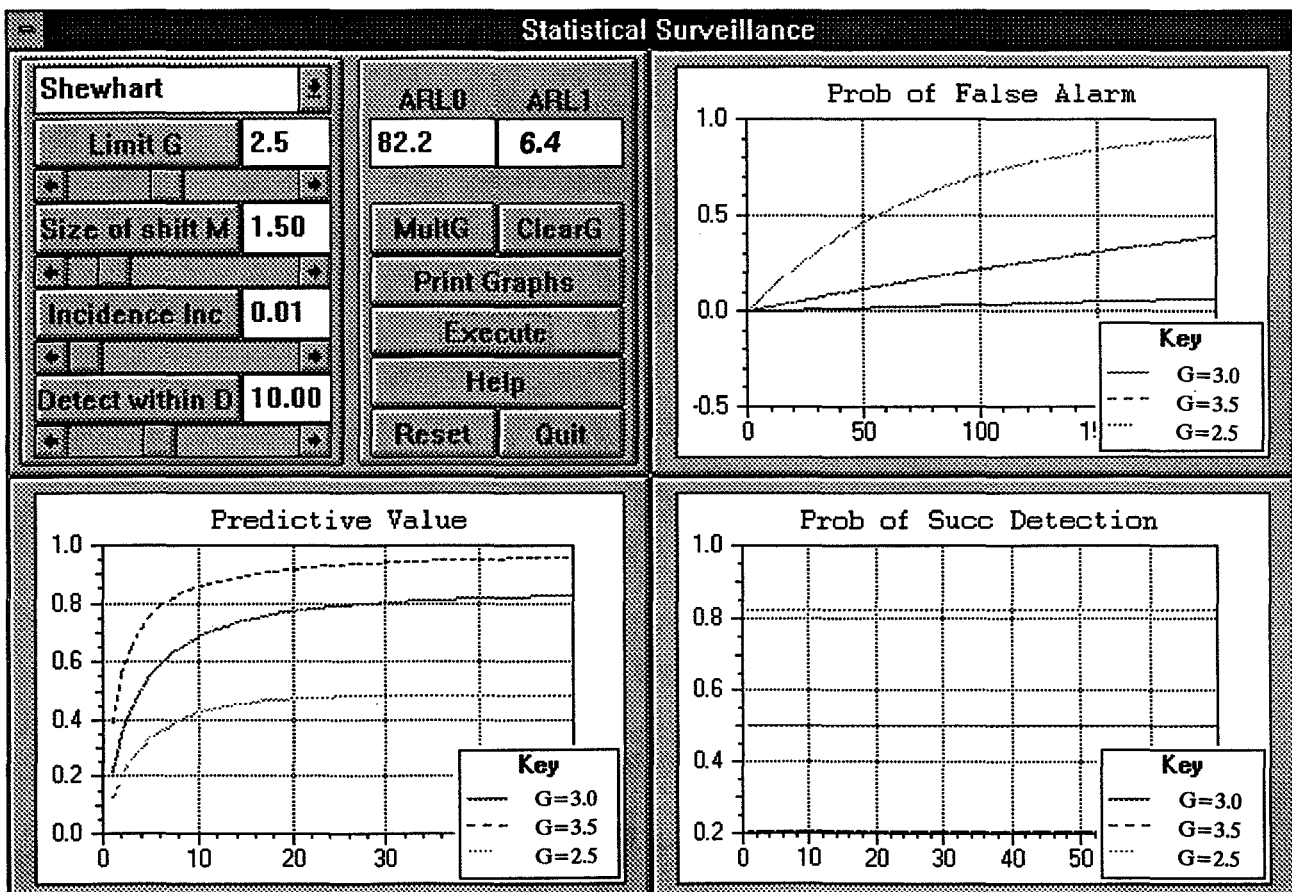


Figure 16. Evaluations of the Shewhart method for different values of G .

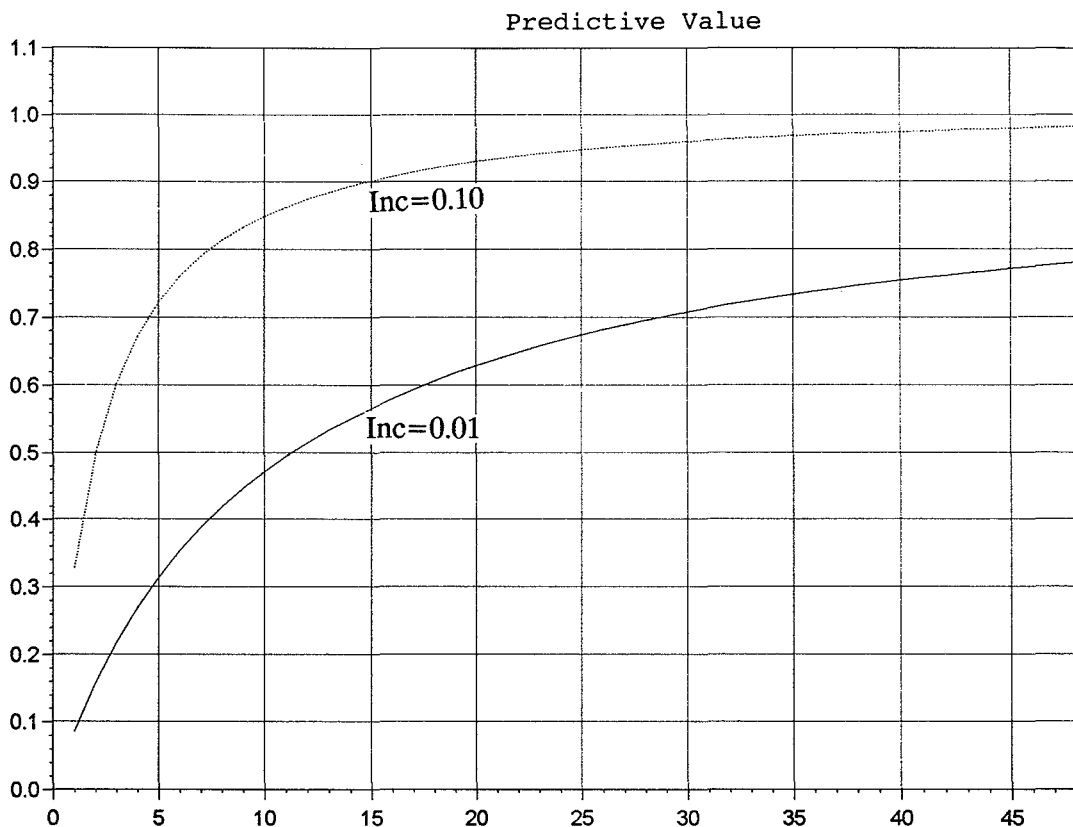


Figure 17. Detailed picture of the predictive value of the Shewhart method for different incidences. $M=1$, $G=3$.

The comparison between the performance of the CUSUM and the Shewhart methods in Figure 15 is made for a case where the two methods have the same ARL_0 and also the same ARL_1 .

By trying different values of a parameter for fixed values of M , Inc and D , as in Figure 16, you get help to choose the right parameter for your application. By varying one variable in a time you will also get insight in how this variable influences the different characteristics. The print function allows detailed prints of each graph as in Figure 17.

ACKNOWLEDGEMENT

This work has been supported by the Swedish Council for Research in the Humanities and Social Sciences.

1992:1	Teräsvirta, T., Tjøstheim, D. & Granger, C.W.J.	Aspects of Modelling Nonlinear Time Series
1992:2	Palaszewski, B.	A conditional stepwise test for deviating parameters
1992::3	Guilbaud, O.	Exact Semiparametric Inference About the Within- Subject Variability in 2 x 2 Crossover Trails
1992:4	Svensson, E. & Holm, S.	Separation of systematic and random errors in ordinal rating scales
1993:1	Frisén, M & Åkermo, G.	Comparison between two methods of surveillance: exponentially weighted moving average vs cusum
1993:2	Jonsson, R.	Exact properties of McNemar's test in small samples.
1993:3	Gellerstedt, M.	Resampling procedures in linear models.
1994:1	Frisén, M.	Statistical surveillance of business cycles.
1994:2	Frisén, M.	Characterization of methods for surveillance by optimality.