



Research Report
Statistical Research Unit
Department of Economics
University of Gothenburg
Sweden

INFERENCE PRINCIPLES FOR MULTIVARIATE SURVEILLANCE

MARIANNE FRISÉN

**Research Report
2011:5
ISSN 0349-8034**

Mailing address:
Statistical Research Unit
P.O. Box 640
SE 405 30 Göteborg
Sweden

Fax
Nat: 031-786 12 74
Int: +46 31 786 12 74

Phone
Nat: 031-786 00 00
Int: +46 31 786 00 00

Home Page:
<http://www.statistics.gu.se/>

INFERENCE PRINCIPLES FOR MULTIVARIATE SURVEILLANCE

MARIANNE FRISÉN

*Statistical Research Unit, Department of Economics, University of Gothenburg,
Gothenburg, Sweden*

ABSTRACT: Multivariate surveillance is of interest in industrial production as it enables the monitoring of several components. Recently there has been an increased interest also in other areas such as detection of bioterrorism, spatial surveillance and transaction strategies in finance. Multivariate counterparts to the univariate Shewhart, EWMA and CUSUM methods have earlier been proposed. A review of general approaches to multivariate surveillance is given with respect to how suggested methods relate to general statistical inference principles. Multivariate on-line surveillance problems can be complex. The sufficiency principle can be of great use to find simplifications without loss of information. We will use this to clarify the structure of some problems. This will be of help to find relevant metrics for evaluations of multivariate surveillance and to find optimal methods. The sufficiency principle will be used to determine efficient methods to combine data from sources with different time lag. Surveillance of spatial data is one example. Illustrations will be given of surveillance of outbreaks of influenza.

Keywords and phrases : Sequential, Surveillance, Multivariate, Sufficiency
AMS (2000) Subject Classification : 62L15, 93C35.

1. INTRODUCTION

The Seventh International Triennial Calcutta Symposium on Probability & Statistics, December 28-31, 2009 was organized jointly by: Department of Statistics, University of Calcutta and Calcutta Statistical Association. The department, established in 1941, with Prof. P.C. Mahalanobis as Head, is the oldest post-graduate Department in Asia offering a course in statistics. It has produced many eminent statisticians who have distinguished themselves both in India and abroad. One of the very distinguished scientists was Samarendra Nath Roy 1906-1964. I was honored to deliver the "S N Roy Memorial Lecture" at the symposium. This paper is based on that talk. The life and work of professor Roy has earlier been described e.g. in a special issue of the Journal of Statistical Planning and Inference. Multivariate analysis has since long been developed in Calcutta. Professor S N Roy developed multiple hypotheses testing and is famous for "Roy's Union-Intersection Principle. His results will here be used for multivariate surveillance.

The aim of surveillance is on-line detection of an important change in an underlying process as soon as possible after the change has occurred. Already at birth surveillance is used, as described by Frisén (2007). The baby might get the umbilical cord around the neck at any time during labor. This will cause a lack of oxygen, and a Caesarean section is urgent. The electrical signal of the heart of the baby during labor is the base for the surveillance system. Detection has to be made as soon as possible to ensure that the baby is rescued without brain damage.

Often several variables are of interest. Multivariate surveillance is of interest in industrial production, for example in order to monitor several sources of variation in assembled products. Wärmefjord (2004) described the multivariate problem for the assembly process of the Saab automobile. Tsung et al. (2008) described the need for multivariate control charts at manufacturing and service processes. The first versions of modern control charts Shewhart (1931) were made for industrial use. Surveillance of several parameters (such as the mean and the variance) of a distribution is multivariate surveillance (see for example Knoth and Schmid (2002)). A common way is dealing with both the mean and the variance is to use Capability index but the theory of multivariate surveillance can suggest alternatives.

In recent years, there has been an increased interest in statistical surveillance also in other areas than industrial production. There is an increased interest in surveillance methodology in the US following the 9/11 terrorist attack. The monitoring of incidences of different diseases and symptoms is carried out by international, national, and local authorities to detect outbreaks of infectious diseases. Transaction strategies based on financial data are of great interest, and in finance, the timeliness of transactions is important. Most theory of stochastic finance is based on the assumption of an efficient market. When the stochastic model is assumed to be completely known, we can use probability theory to calculate the optimal transaction conditions. The support for the efficient market hypothesis depends on the complete knowledge about the model. When the information about the process is incomplete, as for example when a change can occur in the process, there may be an arbitrage opportunity, as demonstrated by Shiryaev (2002). Different aspects of the subject of financial surveillance is described in the book edited by Frisén (2007).

When the collected data involve several related variables, this calls for multivariate surveillance techniques. Spatial surveillance is multivariate since several locations are involved. Multivariate surveillance for financial decision strategies are suggested by for example Okhrin and Schmid (2007) and Golosnoy et al. (2007). The construction of multivariate surveillance methods is based on interesting statistical theory. It also involves practical issues as to the collection of new types of data, and computational ones such as the implementation of automated methods in large scale surveillance data bases. Here the focus

will be on the statistical inference aspects of the multivariate surveillance problem. We will focus on some general approaches for the construction of multivariate control chart methods. These general approaches do not depend on the distributional properties of the process in focus, even though the implementation does. Reviews on multivariate surveillance methods can be found for example in Basseville and Nikiforov (1993), Lowry and Montgomery (1995), Ryan (2000), Woodall and Amiriparian (2002), Frisén (2003), and Sonesson and Frisén (2005).

In Section 2 a review of the univariate case is given. It is demonstrated in Section 3.3 that this is relevant also for multivariate surveillance with simultaneous change. Evaluations in genuinely multivariate cases are treated in Section 3.1. In Section 3.2, different approaches to the construction of multivariate surveillance methods are described and exemplified. Methods for changes with a time lag are treated in Section 3.4. In Section 3.5, we apply a multivariate technique to the detection of influenza outbreaks based on spatial data. Concluding remarks are made in Section 4.

2. UNIVARIATE AND SIMPLE MULTIVARIATE SURVEILLANCE

Complex problems of multivariate surveillance connected will be treated in Section 3. Here we treat the basic problems which are present as soon as on-line surveillance is used in order to detect an important change in the underlying process. This is relevant also for simple multivariate cases as described in Section 3.3.

2.1 Specifications and notations

We denote a univariate process by $X = \{X(t): t = 1, 2, \dots\}$, where $X(t)$ is the observation (vector) made at time t , which here is discrete. The purpose of the monitoring is to detect a possible change, for example the change in distribution of the observations due to the baby's lack of oxygen. The time for the change is denoted by τ . In this section, we consider only one change while more complex cases are treated in Section 3. Before the change, the distribution belongs to the family f^D , and after the time τ , the distribution belongs to the family f^C . At each decision time s , we want to discriminate between two events, $C(s)$ and $D(s)$. For most applications, these can be further specified as $C(s) = \{\tau \leq s\}$ (at the decision time, there has been a change) and $D(s) = \{\tau > s\}$ (at the decision time, no change has occurred yet), respectively. We use the observations $X_s = \{X(t); t \leq s\}$ to form an alarm criterion which, when fulfilled, is an indication that the process is in state $C(s)$, and an alarm is triggered. The change point τ can be regarded either as a random variable or as a deterministic but unknown value, depending on what is most relevant for the application.

The statistical inference in surveillance differs from that of hypothesis testing. The decision concerning whether, for example, the baby is at risk has to be made sequentially, based on the data collected so far. Each new time demands a new decision. Thus, there is no fixed data set but an increasing number of observations. We can never accept any null hypotheses and turn our backs on the mother, since the baby might get the umbilical cord around the neck in the next minute.

2.2 Evaluations

Quick detection and few false alarms are desired properties of methods for surveillance. Evaluations are needed in order to choose which surveillance method to use for a specific aim. Evaluation by significance level, power, specificity, sensitivity, or other well-known metrics may seem convenient. However, these are not easily interpreted in a surveillance situation.

2.3 False alarms

The false alarm tendency is more complicated to control in surveillance than in hypothesis testing since it increases with the number of decisions. The most commonly used measure for surveillance is the Average Run Length when there is no change, $ARL^0 = E(t_A | D)$. A variant of the ARL is the Median Run Length, MRL.

2.4 Delay of the alarm

The most commonly used measure of the delay is the ARL^1 , which is the Average Run Length until the detection of a change (in the situation where the change occurred at the same time as the surveillance started). The part of the definition within parentheses is not always spelled out but generally used.

Shiryayev (1963) suggested that the expected value of the delay from the time of change, $\tau=t$, to the time of alarm, t_A , should be used. It is here denoted by

$$ED(t) = E[\max(0, t_A - t) | \tau=t].$$

The value of $ED(t)$ will typically tend to zero as t increases. Thus, it may be preferred to use the conditional expected delay

$$CED(t) = E[t_A - \tau | t_A \geq \tau = t] = ED(t) / P(t_A \geq t).$$

Sometimes, like when a baby is delivered, the time available for rescuing action is limited. The Probability of Successful Detection, suggested by Frisé (1992), measures the probability of detection with a delay time no longer than a constant d

$$PSD(d, t) = P(t_A - \tau \leq d | t_A \geq \tau = t).$$

2.5 Predicted value

If a method calls an alarm, it is important to know whether this alarm is a strong indication of a change or just a weak one. If τ is regarded as a random variable, this can be done by one summarizing measure. The predictive value of an alarm was suggested by Frisé (1992) as

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

This is the probability that a change has occurred when the surveillance method gives an alarm.

2.6 Optimality

2.6.1 ARL Optimality

Optimality is often stated as a minimal ARL^1 for a fixed ARL^0 . This leads to a concentration on early changes. ARL^1 is the expected run length under the assumption that $\tau=1$ and that the observations at all time points have distributions which belong to f^C . ARL^0 is the expected value of the run length given that no change has occurred and all observations have distributions which belong to f^D . To use efficient methods and evaluate them by the ARL criterion is in conflict with the ancillary principle.

2.6.2 SADT Optimality

An opposite of the ARL criterion is to put emphasis on late changes. The steady state delay time, SADT, measures the delay asymptotically when τ tends to infinity.

2.6.3 Minimal Expected Delay

Since the expected delay depends on τ (the time of the change), it is natural to evaluate by an average of these values $\sum w(t)CED(t)$. If τ is regarded as a random variable, then the weights may be chosen proportional to the density $\pi(t)$ of a change. The average delay is then the expected delay $ED=E(ED(i))$ with respect to the distribution of both τ and t_A . The minimizing of this average, for a fixed false alarm probability, is termed the ED criterion (the minimal expected delay criterion).

2.6.4 Minimax Optimality

A minimax solution, with respect to τ , avoids the requirements of information about the distribution of τ . Moustakides (1986) uses an even more pessimistic criterion, the “worst possible case”, by using not only the least favorable value of the change time, τ , but also the least favorable outcome of $X_{\tau-1}$ before the change occurs. The minimax criterion usually used is the minimum of:

$$\sup_{\tau \geq 1} \sup E_{\tau} \{ [t_A - \tau + 1]^+ | X_{\tau-1} \},$$

for a fixed ARL^0 .

2.7 Methods

Many methods for surveillance can be expressed by a combination of partial likelihood ratios. The likelihood ratio for a fixed value of τ is

$$L(s, t) = f_{X_s}(x_s | \tau=t) / f_{X_s}(x_s | D).$$

The exact formula for these likelihood components will vary between situations. Commonly used methods are often expressed for simple settings like independent Gaussian observations with a step shift. Here, we describe generalized versions by the likelihood expression.

The full likelihood ratio method (LR) is optimal with respect to the criterion of minimal expected delay and also to a wider class of utility functions, as demonstrated by Frisé and de Maré (1991). The alarm set consists of those values of X for which the full likelihood ratio exceeds a limit.

$$t_A = \min \left\{ s; \frac{f_{X_s}(x_s | C(s))}{f_{X_s}(x_s | D(s))} > \frac{P(\tau > s)}{P(\tau \leq s)} \cdot \frac{K}{1-K} \right\},$$

where K is a constant.

The simplest way to aggregate the likelihood components is to add them. Shiryaev (1963) and Roberts (1966) suggested the method, now called the Shiryaev-Roberts method, in which an alarm is triggered at the first time s , for which

$$\sum_{t=1}^s L(s, t) > G,$$

where G is a constant alarm limit.

The method by Shewhart (1931), is simple and the most commonly used method for surveillance. An alarm is given as soon as an observation deviates too much from the target. Thus, only the last observation is considered. The alarm criterion can be expressed by the condition

$$L(s, s) > G,$$

where G is a constant.

The CUSUM method was first suggested by Page (1954) and is reviewed for example in the book by Hawkins and Olwell (1998). The alarm condition of the method can be expressed by the partial likelihood ratios as

$$t_A = \min\{s; \max(L(s, t); t=1, 2, \dots, s) > G\},$$

where G is a constant.

The alarm statistic of the EWMA method is an exponentially weighted moving average,

$$Z_s = (1-\lambda)Z_{s-1} + \lambda X(s), s=1, 2, \dots$$

where $0 < \lambda < 1$ and Z_0 is the target value, which can be normalized to zero.

3. MULTIVARIATE SURVEILLANCE

Multivariate surveillance is of interest in many areas, for example in financial problems, as described by Okhrin and Schmid (2007), and in public health surveillance, as described by Sonesson and Frisén (2005). After discussing evaluations, we will first, in Section 3.2, describe some commonly used stepwise reductions of the multivariate problem and then in Sections 3.3 and 3.4 derive methods from inference principles. An application is given in Section 3.5

3.1 Evaluations

The special problem of evaluation of multivariate surveillance is the topic of the paper by Frisén et al. (2010). In hypothesis testing, the false rejection is considered most important. It is important to control the error in multiple testing since the rejection of a null hypothesis is considered as a proof that the null hypothesis is false. Hochberg and Tamhane (1987) described important methods for controlling the risk of an erroneous rejection in multiple comparison procedures. The False Discover Rate, FDR, suggested by Benjamini and Hochberg (1995) is relevant in situations more like a screening than as hypothesis testing. In surveillance this is further stressed as all methods with a fair power to detect a change have a false alarm rate that tends to one (see Bock (2008)). The problem with adopting FDR is that it uses a probability that is not constant in surveillance. Marshall et al. (2004) solved this problem by monitoring over a short period of time and they use only the properties of the early part of the run length distribution. FDR in surveillance has been advocated for example by Rolka et al. (2007). However, the question is whether control of FDR is necessary when surveillance is used as a screening instrument, which indicates that further examination should be made. Often, the ARL^0 of the combined procedure may be informative enough since it gives information about the expected time until (an erroneous) alarm. It will sometimes be

easier to judge the practical burden with a too low alarm limit by the ARL^0 than by the FDR for that situation.

The detection ability depends on when the change occurs. The conditional expected delay $CED(t) = E[t_A - \tau | t_A \geq \tau = t]$ is a component in many measures, which avoids the dependency on τ either by concentrating on just one value of τ (e.g. one, infinity or the worst value). Frisé (2003) advocated that the whole function of τ should be studied. This measure can be generalized by considering the delay from the first change

$$\tau_{\min} = \min\{\tau_1, \dots, \tau_p\}$$

$$CED(\tau_1, \dots, \tau_p) = E(t_A - \tau_{\min} | t_A \geq \tau_{\min}).$$

The Probability of Successful Detection suggested by Frisé (1992) measures the probability of detection with a delay time shorter than d . In the multivariate case it can be defined as

$$PSD(d, \tau_1, \dots, \tau_p) = P(t_A - \tau_{\min} \leq d | t_A \geq \tau_{\min}).$$

Since the above measures of delay are complex, it is tempting to use the simple ARL measure. The ARL^1 is the most commonly used measure of the detection ability also in the multivariate case. It is assumed that all variables change immediately ($\tau=1$). However, the result in Section 3.3 is that univariate surveillance is always the best method for simultaneous changes. Thus, for genuinely multivariate situations with different change points, ARL^1 is not recommended other as a rough indicator.

3.2 Commonly used approaches

3.2.1 Reduction of Dimension

One way to reduce dimensionality is to consider the principal components instead of the original variables as proposed for example by Jackson (1985), Mastrangelo et al. (1996) and Kourti and MacGregor (1996). In Runger (1996) an alternative transformation, using so-called U^2 statistics, was introduced to allow the practitioner to choose the subspace of interest, and this is used for fault patterns in Runger et al. (2007). Projection pursuit was used by Ngai and Zhang (2001) and Chan and Zhang (2001). Rosolowski and Schmid (2003) use the Mahalanobis distance to reduce the dimensionality of the statistic. After reducing the dimensionality, any of the approaches for multivariate surveillance described below can be used.

3.2.2 Scalar Statistics

The most far going reduction of the dimension is to summarize the components for each time point into one statistic. This is a common way to handle multivariate surveillance problems. Sullivan and Jones (2002) referred to this as “scalar accumulation”. In spatial surveillance it is common to start by a purely spatial analysis for each time point as in Rogerson (1997). A natural reduction is to use the Hotelling T^2 statistic (Hotelling (1947)). Scalars based on regression and other linear weighting are suggested for example by Healy (1987), Kourti and MacGregor (1996) and Lu et al. (1998). Originally, the Hotelling T^2 statistic was used in a Shewhart method, and this is often referred to as the Hotelling T^2 control chart. An alarm is triggered as soon as the statistic $T^2(t)$ is large enough. The reduction to a univariate variable can be followed by univariate monitoring of any kind. Note that, there is no accumulation of information over time of the observation vectors if the Shewhart method is used. In order to achieve a more efficient method, all previous observations should be used in the alarm statistic. There are several suggestions of combinations where reduction to a scalar statistic is

combined with different monitoring methods. Crosier (1988) suggested to first calculate the Hotelling T variable (the square root of $T^2(t)$) and then use this as the variable in a univariate CUSUM method, making it a scalar accumulation method. Liu (1995) used a non-parametric scalar accumulation approach, where the observation vector for a specific time point was reduced to a rank in order to remove the dependency on the distributional properties of the observation vector. Several methods were discussed for the surveillance step, including the CUSUM method. Yeh et al. (2003) suggested a transformation of multivariate data at each time to a distribution percentile, and the EWMA method was suggested for the detection of changes in the mean as well as in the covariance.

3.2.3 Parallel Surveillance

By this commonly used approach, a univariate surveillance method is used for each of the individual components in parallel. This approach can be referred to as combined univariate methods or parallel methods. One can combine the univariate methods into a single surveillance procedure in several ways. The most common is to signal an alarm if any of the univariate methods signals. This is a use of Roy's union-intersection principle for multiple inference problems. Sometimes the Bonferroni method is used to control a false alarm error, see Alt (1985). General references about parallel methods include Woodall and Ncube (1985), Hawkins (1991), Pignatiello and Runger (1990), Yashchin (1994) and Timm (1996).

Parallel methods suitable for different kinds of data have been suggested. Skinner et al. (2003) used a generalized linear model to model independent multivariate Poisson counts. Deviations from the model were monitored with parallel Shewhart methods. In Steiner et al. (1999) binary results were monitored using a parallel method of two individual CUSUM methods. However, to be able to detect also small simultaneous changes in both outcome variables, the method was complemented with a third alternative, which signals an alarm if both individual CUSUM statistics are above a lower alarm limit at the same time. The addition of the combined rule is in the same spirit as the vector accumulation methods presented below. Parallel CUSUM methods were used also by Marshall et al. (2004).

3.2.4 Vector Accumulation

By this approach, the accumulated information on each component is utilised by a transformation of the vector of component-wise alarm statistics into a scalar alarm statistic. An alarm is triggered if this statistic exceeds a limit. This is referred to as "vector accumulation".

Lowry et al. (1992) proposed a multivariate extension of the univariate EWMA method, which is referred to as MEWMA. This method uses a vector of univariate EWMA statistics $\mathbf{Z}(t) = \Lambda \mathbf{X}(t) + (\mathbf{I} - \Lambda) \mathbf{Z}(t-1)$ where $\mathbf{Z}(0) = \mathbf{0}$ and $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$. An alarm is triggered at $t_A = \min\{t; \mathbf{Z}(t)^T \Sigma_{\mathbf{Z}(t)}^{-1} \mathbf{Z}(t) > L\}$ for the alarm limit, L . The MEWMA method can be seen as the Hotelling T^2 control chart applied to EWMA statistics instead of the original data and is thus a vector accumulation method.

One natural way to construct a multivariate version of the CUSUM method would be to proceed as for EWMA and construct the Hotelling T^2 control chart applied to univariate CUSUM statistics for the individual variables. One important feature of such a method is the lower barrier (assuming we are interested in a positive change) of each of the univariate CUSUM statistics. This kind of multivariate CUSUM was suggested by Bodnar and Schmid (2004) and Sonesson and Frisén (2005). Other approaches to construct a multivariate CUSUM have also been suggested. Crosier (1988) suggested the MCUSUM method, and Pignatiello and Runger (1990) had another suggestion. Both these methods use a statistic consisting of univariate CUSUMs for each component and are thus vector accumulation methods. However, the components are used in a different way as compared with the MEWMA

construction. One important feature of these two methods is that the characteristic zero-return of the CUSUM technique is constructed in a way suitable when all the components change at the same time point. However, if all components change at the same time, a univariate reduction is optimal.

3.3 Optimal methods at simultaneous changes

Consider the case where all processes have the same change point so that $\tau_1 = \tau_2 = \dots = \tau_p = \tau$. An example could be when all variables are indicators of the same phenomena. In most evaluations of multivariate surveillance it is assumed that all changes are simultaneous. It now becomes possible to identify the separate factors in the likelihood: the part that depends on the data (but not the value of τ) as well as the part that depends on the s -dimensional vector of partial likelihood ratios. From this it follows that the sequence of the s likelihood ratios is a sufficient sequence. This was proven by Wessman (1998) both for a fixed unknown value of τ and for a stochastic time of change. When the aim is to detect a fully specified, simultaneous change in a multivariate process and the distributions before and after the change are fully specified, it is possible to construct a univariate surveillance procedure based on the sufficient sequence of likelihood ratios. The use of the sufficient statistic implies that no information is lost.

Since a sufficient reduction to univariate surveillance is available, the theory of Section 2 can be applied and optimal methods determined. Healy (1987) derived the CUSUM method for the case of simultaneous change in a specified way for all the variables. The results are univariate CUSUMs for a function of the variables. Since the CUSUM method is minimax optimal, the multivariate methods by Healy (1987) are simultaneously minimax optimal for the specified direction when all variables change at the same time.

3.4 Changes with time lags

We will also consider the case where there are known time lags between the changes of the p processes. There may in some cases be one source of information of good quality that is available after a delay and another source with worse quality that is available early. The multivariate utilization of these data sets might benefit from information on how large the time lag is. Another example is the spatial spread of a disease as will be described in Section 3.5. Sufficient reduction for a step change is derived in Frisén et al. (2011) and for a semiparametric model in Schiöler and Frisén (2010).

3.5 Multivariate outbreak detection

On-line monitoring is needed to detect outbreaks of diseases like influenza. Surveillance is also needed for other kinds of outbreaks, in the sense of an increasing expected value after a constant period. Information on spatial location or other variables might be available and may be utilized. A robust method for outbreak detection by Frisén and Andersson (2009) was adapted to a multivariate case by Schiöler and Frisén (2010).

The relation between the times of the onsets of the outbreaks at different locations was used to determine the sufficient statistic for surveillance. In Schiöler (2011) analyses are made of Swedish influenza data and it is shown that the influenza spreads from the larger cities to the rest of the country with a lag of approximately 1-2 weeks.

The derived maximum likelihood estimator of the outbreak regression was semi-parametric in the sense that the baseline and the slope were non-parametric while the distribution belonged to the one-parameter exponential family. The estimator was used in a generalized likelihood ratio surveillance method. The method was evaluated by Schiöler and Frisén (2010) with respect to robustness and efficiency in a simulation study and applied to spatial data for detection of influenza outbreaks in Sweden .

4. DISCUSSION

Optimality is hard to achieve and even hard to define for all multivariate problems. This is so also in the surveillance case. We have a spectrum of problems where one extreme is that there are hardly any relations between the multiple surveillance components. The other extreme is that we can reduce the problem to a univariate one by considering the relation between the components. Consider, for example, the case when we measure several components of an assembled item. If we restrict our attention to a general change in the factory, changes will be expected to occur for all variables at the same time. Then, the multivariate situation is easily reduced to a univariate one Wessman (1998) and we can easily derive optimal methods. For many applications, however, the specification of one general change is too restrictive. It is important to determine which type of change to focus on. The method derived according to the specification of a general change will not be capable of detecting a change in only one of many components. On the other hand, if we focus on detecting all kinds of changes, the detection ability of the surveillance method for each specific type of change will be small. The more clearly the aim is stated, the better the possibilities of the surveillance to meet this aim. Preferably, the specification should be governed by the application.

The question of which multivariate surveillance method is the best has no simple answer. Different methods are suitable for different problems. Some causes may lead to a simultaneous increase in several variables, and then one should use a reduction to a univariate surveillance method. If the changes occur independently, one does not expect simultaneous changes and may instead prefer to use parallel methods. All knowledge on which component to concentrate on is useful.

The evaluations of multivariate control charts are considerably more complex than for univariate ones. However, the effort to specify the problem is rewarding. Simple measures might be misleading.

REFERENCES

- Alt, F. B. (1985). Multivariate Quality Control, in *Encyclopedia of Statistical Science*, Vol. 6 N. L. Johnson and S. Kotz, eds., pp. 110-122, New York: Wiley.
- Basseville, M. and Nikiforov, I. (1993). *Detection of Abrupt Changes- Theory and Application*, Englewood Cliffs: Prentice Hall.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, *Journal of the Royal Statistical Society B* 57: 289-300.
- Bock, D. (2008). Aspects on the Control of False Alarms in Statistical Surveillance and the Impact on the Return of Financial Decision Systems, *Journal of Applied Statistics* 35: 213-227.
- Bodnar, O. and Schmid, W. (2004). Cusum Control Schemes for Multivariate Time Series. In *Frontiers in Statistical Quality Control. Intelligent Statistical Quality control* Warsaw.
- Chan, L. K. and Zhang, J. (2001). Cumulative Sum Control Charts for the Covariance Matrix, *Statistica Sinica* 11: 767-790.
- Crosier, R. B. (1988). Multivariate Generalizations of Cumulative Sum Quality-Control Schemes, *Technometrics* 30: 291-303.
- Frisén, M. (1992). Evaluations of Methods for Statistical Surveillance, *Statistics in Medicine* 11: 1489-1502.

- Frisén, M. (2003). Statistical Surveillance. Optimality and Methods, *International Statistical Review* 71: 403-434.
- Frisén, M. (2007). *Financial Surveillance*, edited volume, Chichester: Wiley.
- Frisén, M. and Andersson, E. (2009). Semiparametric Surveillance of Outbreaks, *Sequential Analysis* 28: 434-454.
- Frisén, M., Andersson, E., and Schiöler, L. (2010). Evaluation of Multivariate Surveillance, *Journal of Applied Statistics* 37: 2089-2100.
- Frisén, M., Andersson, E., and Schiöler, L. (2011). Sufficient Reduction in Multivariate Surveillance, *Communications in Statistics -Theory and Methods*: to appear.
- Frisén, M. and de Maré, J. (1991). Optimal Surveillance, *Biometrika* 78: 271-80.
- Golosnoy, V., Schmid, W., and Okhrin, I. (2007). Sequential Monitoring of Optimal Portfolioweights, in *Financial Surveillance*, M. Frisé, ed., pp. 179-210, Chichester: Wiley.
- Hawkins, D. M. (1991). Multivariate Quality Control Based on Regression-Adjusted Variables, *Technometrics* 33: 61-75.
- Hawkins, D. M. and Olwell, D. H. (1998). *Cumulative Sum Charts and Charting for Quality Improvement*, New York: Springer.
- Healy, J. D. (1987). A Note on Multivariate Cusum Procedures, *Technometrics* 29: 409-412.
- Hochberg, Y. and Tamhane, A. C. (1987). *Multiple Comparison Procedures*, New York: Wiley
- Hotelling, H. (1947). Multivariate Quality Control, in *Techniques of Statistical Analysis*, C. Eisenhart, M. W. Hastay, and W. A. Wallis, eds., New York: McGraw-Hill.
- Jackson, J. E. (1985). Multivariate Quality Control, *Communications in Statistics. Theory and Methods* 14: 2657-2688.
- Knoth, S. and Schmid, W. (2002). Monitoring the Mean and the Variance of a Stationary Process, *Statistica Neerlandica* 56: 77-100.
- Kourti, T. and MacGregor, J. F. (1996). Multivariate SPC Methods for Process and Product Monitoring, *Journal of Quality Technology* 28: 409-428.
- Liu, R. Y. (1995). Control Charts for Multivariate Processes, *Journal of the American Statistical Association* 90: 1380-1387.
- Lowry, C. A. and Montgomery, D. C. (1995). A Review of Multivariate Control Charts, *IIE Transactions* 27: 800-810.
- Lowry, C. A., Woodall, W. H., Champ, C. W., and Rigdon, S. E. (1992). A Multivariate Exponentially Weighted Moving Average Control Chart., *Technometrics* 34: 46-53.
- Lu, X. S., Xie, M., Goh, T. N., and Lai, C. D. (1998). Control Chart for Multivariate Attribute Processes, *International Journal of Production Research* 36: 3477-3489.
- Marshall, C., Best, N., Bottle, A., and Aylin, P. (2004). Statistical Issues in the Prospective Monitoring of Health Outcomes across Multiple Units, *Journal of the Royal Statistical Society A* 167: 541-559.
- Mastrangelo, C. M., Runger, G. C., and Montgomery, D. C. (1996). Statistical Process Monitoring with Principal Components, *Quality and Reliability Engineering International* 12: 203-210.
- Moustakides, G. V. (1986). Optimal Stopping Times for Detecting Changes in Distributions, *Annals of Statistics* 14: 1379-1387.
- Ngai, H. M. and Zhang, J. (2001). Multivariate Cumulative Sum Control Charts Based on Projection Pursuit, *Statistica Sinica* 11: 747-766.
- Okhrin, Y. and Schmid, W. (2007). Surveillance of Univariate and Multivariate Nonlinear Time Series, in *Financial Surveillance*, M. Frisé, ed., pp. 153-177, Chichester: Wiley.
- Page, E. S. (1954). Continuous Inspection Schemes, *Biometrika* 41: 100-114.

- Pignatiello, J. J. and Runger, G. C. (1990). Comparisons of Multivariate Cusum Charts, *Journal of Quality Technology* 22: 173-186.
- Roberts, S. W. (1966). A Comparison of Some Control Chart Procedures, *Technometrics* 8: 411-430.
- Rogerson, P. A. (1997). Surveillance Systems for Monitoring the Development of Spatial Patterns, *Statistics in Medicine* 16: 2081-2093.
- Rolka, H., Burkom, H., Cooper, G. F., Kulldorff, M., Madigan, D., and Wong, W.-K. (2007). Issues in Applied Statistics for Public Health Bioterrorism Surveillance Using Multiple Data Streams: Research Needs, *Statistics in Medicine* 26: 1834-1856.
- Rosolowski, M. and Schmid, W. (2003). Ewma Charts for Monitoring the Mean and the Autocovariances of Stationary Gaussian Processes, *Sequential Analysis* 22: 257-285.
- Runger, G. C. (1996). Projections and the U^2 Chart for Multivariate Statistical Process Control, *Journal of Quality Technology* 28: 313-319.
- Runger, G. C., Barton, R. R., Del Castillo, E., and Woodall, W. H. (2007). Optimal Monitoring of Multivariate Data for Fault Detection", *Journal of Quality Technology* 39: 159-172.
- Ryan, T. P. (2000). *Statistical Methods for Quality Improvement*, New York: John Wiley & Sons.
- Schiöler, L. (2011). Characterization of Influenza Outbreaks in Sweden, *Scandinavian Journal of Infectious Diseases*: to appear.
- Schiöler, L. and Frisén, M. (2010). Multivariate Outbreak Detection. 2010:2, Statistical Research Unit, Department of Economics, University of Gothenburg, Sweden.
- Shewhart, W. A. (1931). *Economic Control of Quality of Manufactured Product*, London: MacMillan and Co.
- Shiryayev, A. N. (1963). On Optimum Methods in Quickest Detection Problems, *Theory of Probability and its Applications* 8: 22-46.
- Shiryayev, A. N. (2002). Quickest Detection Problems in the Technical Analysis of Financial Data, in *Mathematical Finance - Bachelier Congress 2000*, H. Geman, D. Madan, S. Pliska, and T. Vorst, eds., pp. 487-521, Berlin: Springer.
- Skinner, K. R., Montgomery, D. C., and Runger, G. C. (2003). Process Monitoring for Multiple Count Data Using Generalized Linear Model-Based Control Charts, *International Journal of Production Research* 41: 1167-1180.
- Sonesson, C. and Frisén, M. (2005). Multivariate Surveillance, in *Spatial Surveillance for Public Health*, A. Lawson and K. Kleinman, eds., pp. 169-186, New York: Wiley.
- Steiner, S. H., Cook, R. J., and Farewell, V. T. (1999). Monitoring Paired Binary Surgical Outcomes Using Cumulative Sum Charts, *Statistics in Medicine* 18: 69-86.
- Sullivan, J. H. and Jones, L. A. (2002). A Self-Starting Control Chart for Multivariate Individual Observations, *Technometrics* 44: 24-33.
- Timm, N. H. (1996). Multivariate Quality Control Using Finite Intersection Tests, *Journal of Quality Technology* 28: 233-243.
- Tsung, F., Li, Y., and Jin, M. (2008). Statistical Process Control for Multistage Manufacturing and Service Operations: A Review and Some Extensions *International Journal of Services Operations and Informatics* 3: 191-204.
- Wessman, P. (1998). Some Principles for Surveillance Adopted for Multivariate Processes with a Common Change Point, *Communications in Statistics. Theory and Methods* 27: 1143-1161.
- Woodall, W. H. and Amiriparian, S. (2002). On the Economic Design of Multivariate Control Charts, *Communications in Statistics -Theory and Methods* 31: 1665-1673.
- Woodall, W. H. and Ncube, M. M. (1985). Multivariate Cusum Quality Control Procedures, *Technometrics* 27: 285-292.

- Wärmefjord, K. (2004). *Multivariate Quality Control and Diagnosis of Sources of Variation in Assembled Products*, Gothenburg: University of Gothenburg.
- Yashchin, E. (1994). Monitoring Variance Components, *Technometrics* 36: 379-393.
- Yeh, A. B., Lin, D. K. J., Zhou, H. H., and Venkataramani, C. (2003). A Multivariate Exponentially Weighted Moving Average Control Chart for Monitoring Process Variability, *Journal of Applied Statistics* 30: 507-536.

Research Report

2007:1	Andersson, E.:	Effect of dependency in systems for multivariate surveillance.
2007:2	Frisén, M.:	Optimal Sequential Surveillance for Finance, Public Health and other areas.
2007:3	Bock, D.:	Consequences of using the probability of a false alarm as the false alarm measure.
2007:4	Frisén, M.:	Principles for Multivariate Surveillance.
2007:5	Andersson, E., Bock, D. & Frisé, M.:	Modeling influenza incidence for the purpose of on-line monitoring.
2007:6	Bock, D., Andersson, E. & Frisé, M.:	Statistical Surveillance of Epidemics: Peak Detection of Influenza in Sweden.
2007:7	Andersson, E., Kuhlmann, S., Linde, A & Frisé, M.:	Predictions by early indicators of the progress of the influenza in Sweden.
2007:8	Bock, D., Andersson, E. & Frisé, M.:	Similarities and differences between statistical surveillance and certain decision rules in finance.
2007:9	Bock, D.:	Evaluations of likelihood based surveillance of volatility.
2007:10	Bock, D. & Pettersson, K.:	Explorative analysis of spatial aspects on the Swedish influenza data.
2007:11	Frisén, M. & Andersson, E.:	On-line detection of outbreaks.
2007:12	Frisén, M., Andersson, E. & Schiöler, L.:	A non-parametric system for on-line outbreak detection of epidemics.
2007:13	Frisén, M., Andersson, E. & Pettersson, K.:	Estimation of outbreak regression.
2007:14	Pettersson, K.:	Unimodal regression in the two-parameter exponential family with constant dispersion parameter.
2007:15	Pettersson, K. :	On curve estimation under order restrictions

Research Report

2008:1	Frisén, M.	Introduction to financial surveillance.
2008:2	Jonsson, R.	When does Heckman's two-step procedure for censored data work and when does it not?
2008:3	Andersson, E.	Hotelling's T2 Method in Multivariate On-Line Surveillance. On the Delay of an Alarm.
2008:4	Schiöler, L. & Frisé, M.	On statistical surveillance of the performance of fund managers.
2008:5	Schiöler, L.	Explorative analysis of spatial patterns of influenza incidences in Sweden 1999 – 2008.
2008:6	Schiöler, L.	Aspects of Surveillance of Outbreaks.
2008:7	Andersson, E & Frisé, M.	Statistiska varningssystem för hälsorisker
2009:1	Frisé, M., Andersson, E. & Schiöler, L.	Evaluation of Multivariate Surveillance
2009:2	Frisé, M., Andersson, E. & Schiöler, L.	Sufficient Reduction in Multivariate Surveillance
2010:1	Schiöler, L	Modelling the spatial patterns of influenza incidence in Sweden
2010:2	Schiöler, L. & Frisé, M.	Multivariate outbreak detection
2010:3	Jonsson, R.	Relative Efficiency of a Quantile Method for Estimating Parameters in Censored Two-Parameter Weibull Distributions
2010:4	Jonsson, R.	A CUSUM procedure for detection of outbreaks in Poisson distributed medical health events
2011:1	Jonsson, R.	Simple conservative confidence intervals for comparing matched proportions
2011:2	Frisé, M	On multivariate control charts
2011:3	Frisé, M	Methods and evaluations for surveillance in industry, business, finance, and public health
2011:4	Knoth, S & Frisé, M	Minimax Optimality of CUSUM for an Autoregressive Model