



Research Report  
Department of Statistics  
Göteborg University  
Sweden

---

## Detection of environmental catastrophes

Eric Järpe

Research Report 2000:6  
ISSN 0349-8034

---

Mailing address:	Fax	Phone	Home Page:
Dept of Statistics	Nat: 031-773 12 74	Nat: 031-773 10 00	<a href="http://www.stat.gu.se/stat">http://www.stat.gu.se/stat</a>
P.O. Box 660	Int: +46 31 773 12 74	Int: +46 31 773 10 00	
SE 405 30 Göteborg			
Sweden			

# Detection of Environmental Catastrophes

E. Järpe

Department of Statistics, Göteborg University,  
SE 405 30 Göteborg, Sweden  
E-mail: eric.jarpe@statistics.gu.se

## Abstract

For on-line detection of environmental catastrophes spatial as well as temporal effects are crucial. Observations of some substance are made at different locations in space at discrete times. Mainly two spatial features are considered: first, a situation where the observations made at the same time but at different sites are correlated, and secondly, a situation where the catastrophe originates from a source with known position in space and then spreads, passing other known positions as time passes. The first situation turns out to be a special case of a previously studied situation in multivariate surveillance. The second case is reduced to a univariate problem and a brief evaluation of the Shewhart method, the Cusum method and the Shiryaev-Roberts method is made. The theory is applied on the case of surveillance of radiation. The suggested methods are compared with the method presently in use in Sweden.

*Keywords:* Spatial process, Multivariate normal distribution, Simply ordered shift process, Radiation data.

## 1 Introduction

In several environmental issues, such as radiation change detection, forestry disease surveillance, earthquake warning system, climate change detection and others, there is a need for methods to judge whether a change in a spatial process has occurred or not. It is crucial to know if strategies of evacuation, treatment, public warnings, preventive actions etc. might have to be taken. Previous studies are made by e.g. Rogerson (1997), who considered an index (the Tango statistic) to detect a change in clustering, and Järpe (1999, 2000), who considered change of the parameter for spatial interaction, of the Ising

model and in an Ising dynamic model. Several studies on environmental issues in spatial data were treated by Lawson (see e.g. Lawson et al, 1999).

Suppose that we want to detect whether a nuclear incident occurs within or outside a geographical region. We measure radiation at stations located inside the geographic region and we assume that an incident could be recognised as an increased level of radiation. We could also consider the situation where we want to react to a change in pollution in a lake or a river maybe near some chemical industry. Then, observations of the polluting substance's concentration might be made. In this paper we only consider the detection of a nuclear incident in Sweden or in a nearby country in detail but the list of situations which could benefit from the results derived here could probably be made much longer. In Sweden there are 37 stations, fairly evenly spread out across the country, measuring gamma radiation. They are administrated by the Swedish Radiation Protection Institute responsible for determining whether an incident has occurred or not. If an incident does occur it might be recognised as a spreading disc of increased radiation levels. Thus, concentric spread around a source in the lattice is one possible scenario. Other parameters such as wind, topography, distance to the source etc. might contribute to modified scenarios.

In Section 2 we describe the spatial model and the shift process considered here. The spatial model is a multivariate normal one with a covariance matrix that has a spatial interpretation (measurements made at nearby stations would presumably be more similar than measurements made farther apart). However, for the main part of this paper we will assume no spatial correlations but rather concentrate upon a spatial shift process, i.e. the shift spreads from a source and through the geographic region depending on the time of shift at the source and the distances between the stations at which observations are made, and the source. When it comes to the features of environmental issues, wind speed, direction or water current may play an important role. Many spatial and spatio-temporal surveillance investigations do not take this into account.

In Section 3 we consider some surveillance methods to detect a shift in the mean. In general there might be a change in the expectation, covariance or the variance of the process. We consider changes in marginal expectations from one constant to another constant. First the case when the shift occurs at the same time at certain sites and nowhere else, with spatial correlations included in the model, is considered. Secondly the case when the shift occurs at different times depending on wherefrom the shift derives and how it spreads, without spatial correlations in the model, is considered.

An example of surveillance of radiation is given in Section 4. It is based upon gamma radiation data from 5 stations administered by the Swedish

radiation protection institute. Having calibrated the observed false alarm rate, delay times of the first motivated alarm are given. The alarm functions are plotted to give illustrations of how the method used today performs for this data compared to some other methods.

Finally, in Section 5, the results and further research are discussed.

## 2 Spatial Model

In order to detect a change from the process in control (i.e. before a catastrophe has occurred) it is necessary to model how covariates influence it. For the rest of this section and in Section 3, we shall assume that this analysis has been made and only be dealing with observations of the residuals from a known model of the process in control. In Section 4, the transformation to these residuals is illustrated.

Observations of a random process,  $X = \{X_i(t) : i \in A_N, t \in \mathbb{Z}_+\}$  (called the spatial process), of  $N \times 1$  vectors,  $X(t)$ , are made at locations (called sites), in a region  $R \subset \mathbb{R}^2$ , denoted by  $1, 2, \dots, N$ , and the set of sites,  $\{1, 2, \dots, N\}$  is called  $A_N$ . The set of all positive integers is denoted by  $\mathbb{Z}_+$ , and  $\mathbb{R}$  is the set of all real numbers. The index  $t \in \mathbb{Z}_+$  symbolises equidistant times at which observations of  $X(t)$  are made. Furthermore, let us assume that there is also a source in  $R$ , the location of which is denoted by  $k$ . Observations of  $X(t)$  will be denoted by  $x(t)$  and observations of  $X_{\leq s} = \{X(t) : 1 \leq t \leq s\}$  by  $x_{\leq s}$ .

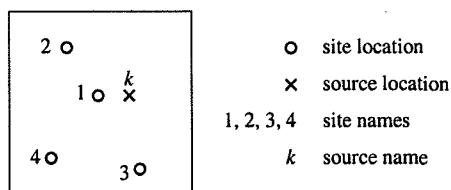


Figure 1: *The lattice consists of  $N$  sites (in this illustration  $N = 4$ ) and one source, all located in the study region.*

Let  $\tau$  be the random time-point of a catastrophe. Let  $\mu = \{\mu_i(t) : i \in A_N, t \in \mathbb{Z}_+\}$  be a shift process which, given  $\tau$ , is a deterministic function of  $\tau$ ,  $\Sigma$  be a covariance matrix which has a spatial interpretation in the sense that the covariance between two sites is inversely proportional to the distance between these sites (assuming unit variances,  $\Sigma_{ii} = 1$ , means no restriction, see further Section 2.2), and  $\epsilon = \{\epsilon_i(t) : i \in A_N, t \in \mathbb{Z}_+\}$  be a noise process

of  $N \times 1$  vectors with, at all levels, independent standard normal random variables. We assume that the spatial process  $X$  can be represented as

$$X = \mu + \Sigma^{1/2}\epsilon.$$

Adding covariates (such as climate effects),  $V = \{V_i(t) : i \in A_N, t \in \mathbb{Z}_+\}$  the values of which are assumed to be given, the “true” observation process,  $\Xi = \{\Xi(t) : t \in \mathbb{Z}_+\}$ , may be represented as a function of  $V$  and  $X$ , and observations of  $\Xi$  will be denoted by  $\xi$ .

The expectation shift process, a deterministic function of the random variable  $\tau$ , indicates that for  $t < \tau$ ,  $E[X_i(t)] = 0$  (no restriction) and for  $t \geq \tau$ ,  $E[X_i(t)] = \delta_i$  (see Section 2.1).

## 2.1 Shift Process

The shift may spread in many different ways. We consider types of shift processes where the shift starts spreading at random time  $\tau$  reaching site  $i$  at time  $\tau + u_i$  where  $u_i$  is a site  $i$  specific integer time distance. The only restrictions are that the random time  $\tau$ , when the shift starts, completely determines both  $u_i$  and the order of changing sites and that each site having changed cannot change back; thus the sequence of sets of sites that have changed must be non-decreasing in  $t$ . In many applications the shift size would decay as the distance to the source increases. However, for this first attempt to consider the problem of a spreading shift, we assume known constant levels, 0 and  $\delta_i$ ,  $i \in A_N$ . This means that for each site  $i$ , the expectation  $E[X_i(s)]$  shifts from 0,  $s < \tau + u_i$  to  $\delta_i$ ,  $s \geq \tau + u_i$  (since we can always define  $x(s)$  by  $\xi_i(s) - v_i(s) - E[X_i(s)|\tau > s]$ ). The derivations in Section 3 are made for the case of a positive global shift of size  $\delta_i = \delta > 0$  for each  $i \in A_N$ .

**Definition 1** *A simply ordered shift process is a sequence of expectation parameter vectors  $\{\mu(s) : s \in \mathbb{Z}_+\}$  with components  $\mu_i(s) = \delta I(\tau \leq s - u_i)$  such that  $\{u_i : i \in A_N\}$  is simply ordered with respect to  $i$ .*

In this definition,  $I(\cdot)$  denotes the indicator function where  $I(A)$  equals 1 whenever  $A$  is true and 0 otherwise. Let us look at a few examples.

A simple situation is **no spread**: at time  $\tau$  there is a shift in the mean at sites  $i \in B \subset A_N$  that remain shifted and the other sites  $i \in A_N \setminus B$  remain unshifted (see Figure 2). In other words site  $i$  shifts at time  $\tau + u_i$  where

$$u_i = \begin{cases} 0 & \text{if } i \in B \\ \infty & \text{if } i \notin B. \end{cases}$$

$B$  may be either known or random. One special case of this is when there is a shift at all sites (i.e.  $B = A_N$ ). Another is when there is a shift at only one site. In the latter case it is meaningful to further define whether this one site is known or random in  $A_N$ . Variants of this are e.g. the cases when only one site shifts unknown which, or when a certain known number of sites shift but which sites is not known.

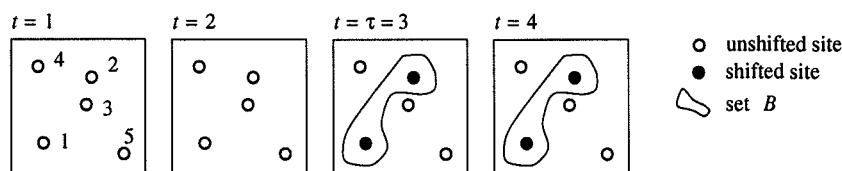


Figure 2: *No spread: there is only one time of shift. The sites in  $B$  shift at time  $\tau$  and the others never shift (i.e. shift at time  $t = \infty$ ).*

Let us next consider the situation where the shift spreads in space (Figure 3). At a random time  $\tau$  there is an incident at location  $k$  which causes a cloud to grow concentrically around  $k$  thus including more and more sites as time passes. With the *integer part* defined as  $\lfloor y \rfloor = \max\{m \in \mathbb{Z} : m \leq y\}$  the time of shift at site  $i$  is  $\tau + u_i$  where  $u_i = \lfloor a^{-1}d(i, k) \rfloor$ ,  $a$  is a speed-of-spread specific constant and  $d(l_1, l_2)$  is the Euclidian distance between two locations,  $l_1$  and  $l_2$ , in  $R$ . This kind of spread could be adequate in case of calm weather and it will be called **concentric spread**.

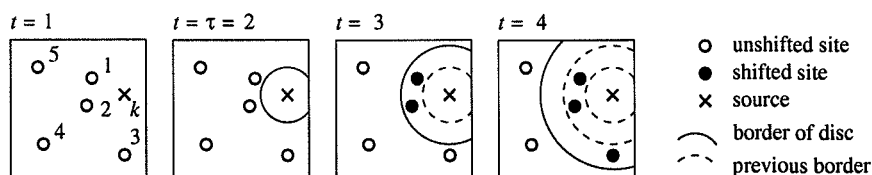


Figure 3: *Concentric spread: the shift spreads concentrically from a source and reaches sites successively as time passes.*

A variant is **concentric spread with a drift** (Figure 4). At time  $\tau$  a cloud starts spreading from the source, at  $k(0)$ , and at each time following  $t = \tau + 1, \tau + 2, \dots$  the centers of the disc shaped clouds' locations are  $k(1), k(2), \dots$  in such a way that  $d(k(t+1), k(t)) \leq a$  for  $t = 1, 2, \dots$ , which might be a reasonable assumption in case of a mild wind. This makes the subsets of sites an increasing sequence and therefore the shift process a simply ordered one. Thus, similarly to concentric spread,  $u_i = \lfloor a^{-1}d(i, k(t-\tau)) \rfloor$ . Concentric

spread with a drift could perhaps be reasonable in case of a mild wind or water current, moving the spreading cloud in the wind or current direction.

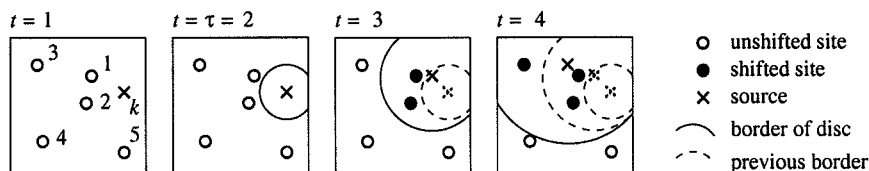


Figure 4: *Concentric spread with a drift: the shift spreads according to a moving and increasing disc which reaches sites successively as time passes. Observe that the discs must “include each other” in order for the shift process to be simply ordered.*

Another variant is **sector spread** (Figure 5) where sites shift in the same way as in the concentric spread but are restricted to a certain sector.

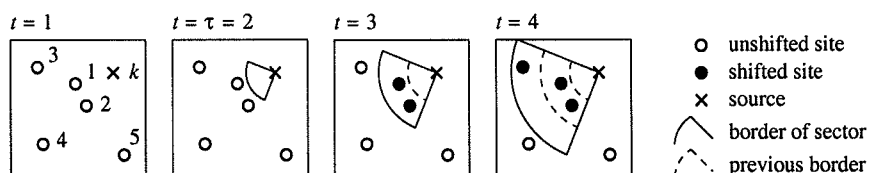


Figure 5: *Sector spread: the shift spreads concentrically restricted to a sector. The sites within the sector shift eventually while, in this example, site 5 never shifts.*

Having introduced a Cartesian plane this sector may be given by two angles,  $\alpha_1$  and  $\alpha_2$ , possibly reflecting e.g. topographic structure. Thus

$$u_i = \begin{cases} \lfloor a^{-1}d(i, k) \rfloor & \text{if } \alpha_1 \leq \arg(\mathbf{a}_i) \leq \alpha_2 \\ \infty & \text{otherwise} \end{cases}$$

where  $\mathbf{a}_i$  is the coordinate pair of site  $i$  in the Cartesian plane with origin in  $k$ . (Observe that  $\arg(\mathbf{a}_i) \notin [\alpha_1, \alpha_2]$  means that the mean at site  $i$  never changes.) This might be applicable when the wind or water current is strong and the source is close to the sites.

If the source is very far away from the sites, the border of the concentric spread will be close to a straight line (Figure 6). Introducing a Cartesian plane with origin in  $k$  and one axis (let us call it  $y$ -axis) parallel to the spread line, the shift could be modelled as occurring when passed by a line starting through  $k$  moving in the direction of the  $x$ -axis, orthogonal to the

$y$ -axis. Thus site  $i$  shifts at time  $\tau + u_i$  where  $u_i = \tau + \lfloor a^{-1}i_x \rfloor$  and  $i_x$  is the  $x$ -coordinate of  $i$ . This will be called **line spread**.

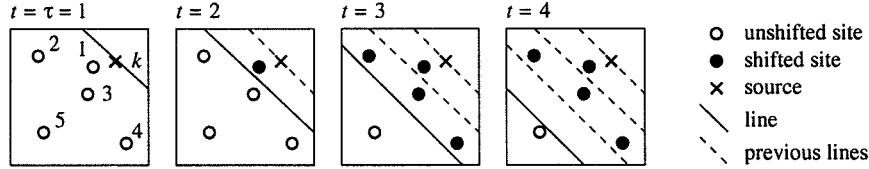


Figure 6: *Line spread: the shift spreads as a line moving across the study region.*

These are a few scenarios according to simply ordered shift processes but of course one could think of many more.

Consider the special case of a simply ordered shift process when there is exactly one site at each time interval.

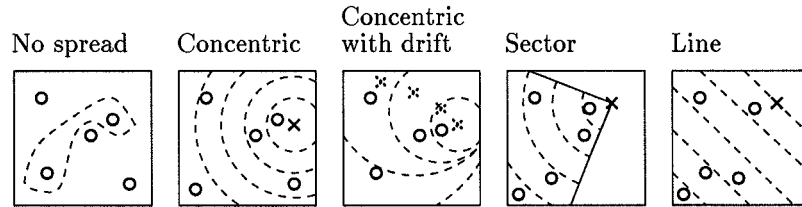


Figure 7: *With certain spatial structures, such that there is exactly one site in each new piece added to the growing cloud, all of the spreading scenarios mentioned above, except for the first, “no spread”, are examples of simply ordered shift processes with equidistant sites.*

This may (after a renumbering of the sites) be modelled as  $u_i = i$  and it is a restriction on the spatial structure. A **simply ordered shift process for equidistant sites** is a simply ordered shift process with components  $u_i = i$ . Some lattice structures are examples of a simply ordered shift process with equidistant sites as illustrated in Figure 7.

## 2.2 Spatial Covariances

The covariance matrix  $\Sigma$  considered consists of components of the form

$$\Sigma_{ij} = \begin{cases} \phi/d(i,j)^2 & \text{when } i \neq j \\ 1 & \text{otherwise} \end{cases}$$



where  $\phi$  is a covariance magnitude parameter. This may also be written as  $I_N + \phi D_N$  where  $I_N$  is the  $N \times N$  identity matrix and  $D_N$  is the  $N \times N$  matrix with  $\text{diag}(D_N) = \mathbf{0}$  and entries  $d(i, j)^{-2}$  elsewhere. (Observe that it means no restriction to assume that  $\text{Var}[X_i(t)] = 1$  instead of  $\sigma^2$  since it just implies a scaling of the covariance parameter  $\phi$  and shift size parameter  $\delta$ .)

If instead the covariance matrix  $\Sigma$  was defined so that the diagonal of  $I_N - \Sigma^{-1/2}$  was null then, for each fixed  $t$ ,  $X(t)$  would be a spatial autoregressive process (Whittle, 1954). Conditionally on  $\tau$ , this may be written as

$$X_i(t) = \sum_{j \neq i} C_{ij} X_j(t) + \mu'_i(t) + \epsilon_i(t) \quad \text{for } i \in A_N$$

with  $C = I_N - \Sigma^{-1/2}$ ,  $\mu'(t) = \Sigma^{-1/2} \mu(t)$  constants (since  $\tau$  is given) and  $\epsilon_i(t) \sim N(0, 1)$ ,  $\epsilon_i(t)$  and  $\epsilon_j(t)$  independent,  $i, j \in A_N : i \neq j$ . Nevertheless we will not make this assumption but rather let  $\Sigma = I_N + \phi D_N$  as previously defined which fully specifies all covariances.

Cressie (1993) studies more sophisticated models for the covariance matrix.

### 3 Surveillance

At some time and place a source starts causing shifts that may spread. The surveillance problem is to detect the change accurately and quickly. For this purpose one might consider different kinds of in-control events,  $D(s)$ , and out-of-control events,  $C(s)$ . Here, it is relevant with  $D(s) = \{\tau > s\}$  and  $C(s)$  some subset of  $\{\tau \leq s\}$  where  $s$  is the decision time minus the number time intervals between the source and the site closest to the source,  $\min_i u_i$ , which is denoted by  $u_{(1)}$ .

There are several general surveillance methods suggested in the literature (see e.g. Frisén (1999) or Lai (1995)). Here we give examples of how the Shewhart, the Cusum and the Shiryaev-Roberts (SR) methods turn out for the situations considered.

In this paper we let  $D(s) = \{\tau > s - u_{(1)}\}$  and  $C(s) = \{\tau \leq s - u_{(1)}\}$ . The partition  $\{C'(1), C'(2), \dots, C'(s)\}$  of  $C(s)$  where  $C'(t) = \{\tau = t - u_{(1)}\}$  will also be used. Let  $\{x(t) : t \leq s\}$  be denoted by  $x_{\leq s}$  and the corresponding random variables by  $X_{\leq s}$ .

All three methods studied here are based on combinations of the partial likelihood ratios

$$L(s, t) = \frac{f(x_{\leq s} | C'(t))}{f(x_{\leq s} | D(s))}.$$

The stopping rules can be expressed as

$$t_A = \inf\{s \geq 1 : p(\{L(s,t) : 1 \leq t \leq s\}) > c\}$$

where  $p$  is an alarm function and  $c$  is a chosen constant (sometimes called “threshold”).

The **Shewhart method** for the full likelihood ratio, is here defined as the stopping rule with alarm function

$$p(x_{\leq s}) = L(s, s).$$

A Shewhart stopping rule for the multivariate situation may be defined in different ways. Originally Shewhart defined it as the first time at which an observation exceeds a threshold (Shewhart, 1931). In the multivariate case this should be a method based on the vector,  $x(s) = [x_1(s) \ x_2(s) \ \dots \ x_N(s)]$ , of observations made at the last time-point,  $s$ . The information from the  $N$  components can be combined in different ways. We study the union-intersection, the nearest location and the full likelihood ratio techniques. The nearest location principle and the full likelihood ratio results in the alarm function  $p(x_{\leq s}) = x_1(s)$  (assuming site 1 is closest to the source) which is minimal sufficient for discriminating  $\{\tau = s - u_{(1)}\}$  from  $\{\tau > s - u_{(1)}\}$ .

The **Cusum method** was suggested by Page (1954) and further investigated by e.g. Moustakides (1986). It has the alarm function

$$p(x_{\leq s}) = \max_{1 \leq t \leq s} L(s, t).$$

The **Shiryaev-Roberts method**, studied by Shiryaev (1963) and Roberts (1966), has the alarm function

$$p(x_{\leq s}) = \sum_{t=1}^s L(s, t).$$

We will now apply the general methods to the spatial models described in Section 2.

### 3.1 No Spread but Spatial Covariances

With “no spread” we mean that all shifts occur at the same time,  $\tau$ . Consider the case defined in Section 2 with

$$u_i = \begin{cases} 0 & \text{if } i \in B \\ \infty & \text{if } i \notin B. \end{cases}$$

$B$  may be a known subset of  $A_N$  or a random variable  $B : \Omega_B \rightarrow \mathcal{B}_N$  where  $\mathcal{B}_N \subseteq \sigma(A_N)$ , a subset of the smallest set containing all subsets of  $A_N$ .

### 3.1.1 Known Shift Region

When the set  $B$  is known, there is a simple solution. Wessman (1998) proved that there is a simple minimal sufficient statistic for discrimination between  $C(s)$  and  $D(s)$  for this case.

Let  $M$  be the statistic

$$M(X(s)) = Z^{-1/2} \sum_{i \in A_N} w_i X_i(s)$$

where  $Z = \sum_{i \in B} w_i$ ,  $w_i = \sum_{j \in B} \Sigma_{ij}^{-1}$  and  $\Sigma_{ij}^{-1}$  is the  $i$ th row  $j$ th column component of  $\Sigma^{-1}$ . Then  $M(X(s))$  is sufficient for discrimination between  $C(s)$  and  $D(s)$ .

Given  $\tau = t$ ,

$$M(X(s)) \sim \begin{cases} N(0, 1) & \text{if } s < t \\ N(\delta Z^{1/2}, 1) & \text{if } s \geq t \end{cases}$$

and thus the surveillance is a special case of univariate surveillance which is studied by e.g. Frisén (1999), Frisén and Wessman (1999) and Järpe and Wessman (1999).

**Example** Suppose  $N = 2$ ,  $B = \{1\}$  and  $\text{Cov}[X_1(t), X_2(t)] = \rho$ . Then

$$\Sigma^{-1} = (1 - \rho^2)^{-1} \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix} \quad \text{so } Z = (1 - \rho^2)^{-1}$$

and thus

$$M(X(s)) = (1 - \rho^2)^{-1/2} (X_1(s) - \rho X_2(s)).$$

We get for instance the Shewhart method as

$$t_A = \min\{s : X_1(s) - \rho X_2(s) > c_1\}$$

and the Cusum method as

$$t_A = \min\{s : \max_{1 \leq t \leq s} \sum_{r=t}^s (X_1(r) - \rho X_2(r) - \delta/2) > c_2\}.$$

### 3.1.2 Unknown Shift Region

Sometimes the probability distribution of the shift region,  $B$ , might be known. Then, the likelihood ratio is

$$\begin{aligned} \log \frac{f(x(s)|\tau \leq s)}{f(x(s)|\tau > s)} &= \log \sum_{b \subseteq B_N} P[B=b] \frac{f(x(s)|\tau \leq s, B=b)}{f(x(s)|\tau > s, B=b)} \\ &= \log \sum_{b \subseteq B_N} P[B=b] \exp\left(\delta \sum_{i \in A_N} (X_i(s) - w_i)\right) \end{aligned}$$

where  $w_i = \frac{\delta}{2} \sum_{j \in b} \Sigma_{ij}^{-1}$ . Its distribution is not purely Gaussian as in the case with known  $B$ . Either  $B$  may take its values in all subsets of  $A_N$  or only subsets of a certain size. In the latter situation a special case is that the size is 1 and the sample space of  $B$  is then  $\{1, 2, \dots, N\}$ .

Often, however, no information on  $P[B=b]$  is available. Then we have to rely on general results on multivariate surveillance. Wessman (1999) made a very thorough study of the situation with  $N=2$  and with  $\mathcal{B}_N$  as the smallest sigma algebra of  $A_N$ :  $\{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$  (or  $\{(00), (10), (01), (11)\}$  in Wessman's notation). Wessman examined different ways of weighing together the information from the last observation made at the two sites. Thus he considered methods of Shewhart type, in the sense "methods that only take the last observation into account": two union-intersection methods based on the marginal distributions of  $X_i(s)$ ,  $i = 1, \dots, N$  and three methods based on the joint distribution of  $X(s)$  (Hawkins method, Hotellings  $T^2$ -method, a likelihood ratio method). For these methods Wessman found their respective alarm regions. He also derived their properties in respect of several common optimality criteria such as alarm probability, probability to detect the event that the first change occurs at time  $t$ ,  $P[\tau = t, B \text{ is non-empty}]$ , probability of successful detection, conditional expected time to alarm  $E[t_A | \tau = 1, B = \{1\}]$  and  $E[t_A | \tau = 1, B = \{1, 2\}]$ . He also considered the dependence between the change-points of the different variables. In this paper, given the change-point of one of the sites, all change-points are given.

## 3.2 Simply Ordered Shift Process and Spatial Independence

### 3.2.1 A Union-intersection Technique

The union-intersection principle (Roy, 1953) is a commonly used approach in multivariate data analysis as a way of utilising information from different processes without much knowledge of their relations. Applying this principle to stopping rules means simply to combine several marginal stopping rules, say  $t_1, \dots, t_n$ , and stop as soon as any one of them stops,  $t = \min_i t_i$ . We will consider the following union-intersection technique based on the alarm function,  $p(X_{i, \leq s})$ ,

$$t_A = \min\{s \geq 1 : \max_{i \in A_N} p(X_{i, \leq s}) > c\}.$$

A union-intersection statistic based on the marginal distributions might make sense for a situation where the marginal processes are to be treated equally with respect to time. But if a spatial shift deriving from a source

is considered, distances to the source will make some sites shift before others. Therefore a union-intersection technique for this case can not be expected to be as efficient as techniques that can utilise full information. This union-intersection technique is applied and compared to other techniques in Section 4.

### 3.2.2 Methods Utilising Knowledge About the Shift Process

If the source (i.e. its position) is known and the spread is concentric then of course the design with all sites as close as possible to the source would result in maximum power and reliability of the surveillance methods. Thus, since all stations would be at equal minimal distance from the source, the surveillance problem would be that treated in Section 3.1 with the set  $B$  of sites that shift equal to  $\{1, 2, \dots, N\}$ .

However, with wind in the model we would not be able to tell in advance the order of the shift process. We would be able though to register wind direction and speed at each time  $t$  and (assuming the wind direction and speed to be homogenous for the whole study area) given these data calculate how an imaginary cloud would drift.

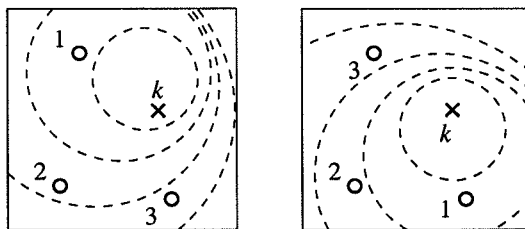


Figure 8: *Concentric spread with a drift. The direction and speed of the drift determines the order of the shift process.*

Nevertheless the spatial structure might be “almost equidistant”, i.e. in most time intervals there is exactly one site but in a few intervals there are more than one site or no site at all.

We begin by defining the diagonal process in the general (meaning not necessarily equidistant) spatial structure. Let us transform the spatial process

$$\begin{aligned} X &= \{X_1(1), \dots, X_N(1), X_1(2), \dots, X_N(2), \dots\} \\ &= \{X_i(t) : i \in A_N, t \in \mathbb{Z}_+\} \end{aligned}$$

into

$$\begin{aligned} Y &= \{Y(1,1), Y(2,1), Y(2,2), Y(3,1), Y(3,2), Y(3,3), \dots\} \\ &= \{Y(s,t) : s \in \mathbb{Z}_+, t=1, \dots, s\} \end{aligned}$$

where the process  $Y$  is defined as follows.

**Definition 2** *Let the diagonal process  $Y$  be defined by*

$$Y = \{Y_{\leq s} : s \in \mathbb{Z}_+\}$$

where

$$Y_{\leq s} = \{Y(s,t) : t=1, \dots, s\}, \quad Y(s,t) = n_t^{-1} \sum_{r=1}^{n'_t} \sum_{i:u_i=r} X_i(t+r-1),$$

$$n_t = \#\{i : u_i \leq s-t+1\} \text{ and } n'_t = \min(\max_i u_i, s-t+1) - \min_i u_i + 1.$$

**Proposition 1** *The statistic  $Y_{\leq s}$  is minimal sufficient for discriminating between  $C(s) = \{\tau \leq s\}$  and  $D(s) = \{\tau > s\}$ .*

(A proof of this is given in Appendix A.)

Let  $B_0$  denote the unique spatial structure configuration of Section 3.2 where  $u_i = i$  for all  $i \in \mathbb{Z}_+$  meaning that there is exactly one site in each unit interval.

Let  $b_1$  be a spatial structure which is the same as  $B_0$  except at one interval where there are either two sites or none. In this sense  $b_1$  differs from  $B_0$  by exactly one site. Also let  $B_1$  denote the family of spatial structures that differ from  $B_0$  by exactly one site. In general let  $B_m$  denote the family of spatial structures that differ from  $B_0$  by exactly  $m$  sites. The smaller  $m$  is, the better the properties of surveillance of the model in Section 3.2 with equidistant sites can often be expected to approximate the properties of the variables of the process for a spatial structure of  $B_m$ .

**Proposition 2** *In a spatial structure of  $B_m$*

$$Y(s,t) \sim \begin{cases} N(0, v_t) & t < \tau \\ N(\delta, v_t) & t \geq \tau \end{cases}$$

where  $(\min(N, s-t+1) + m)^{-1} \leq v_t \leq (\min(N, s-t+1) - m')^{-1}$  and  $m' = \min(m, N, s-t+1)$ .

(A proof of this is given in Appendix A.)

The surveillance methods in the simply ordered case having relaxed the equidistance restriction are constructed essentially in the same way from the diagonal process  $Y$ ; the only difference is how  $Y$  is constructed. The Shewhart method in this special case is

$$t_A = \min\{s \geq 1 : y(s, s) > \delta^{-1} v_{s-u(1)} \log c - \delta/2\}.$$

The Cusum method is

$$t_A = \min\left\{s \geq 1 : \max_{1 \leq t \leq s-u(1)} \sum_{r=t}^{s-u(1)} v_r^{-1} (y(s, r) - \delta/2) > \delta^{-1} \log c\right\}.$$

The Shiryaev-Roberts method is

$$t_A = \min\left\{s \geq 1 : \sum_{t=1}^{s-u(1)} \exp\left(\delta \sum_{r=t}^{s-u(1)} v_r^{-1} (y(s, r) - \delta/2)\right) > c\right\}.$$

### The Equidistant Case

When the sites are equidistant with respect to the shift process, the smallest distance to the source is  $u(1) = \min_i u_i = 1$ . (Throughout the rest of the paper we make use of the convention that  $\sum_{a=a_0}^{a_1} f(a) = 0$  whenever  $a_1 < a_0$ .) In this special case, the diagonal process is  $Y = \{Y_{\leq s} : s \in \mathbb{Z}_+\}$  where

$$Y_{\leq s} = \{Y(s, t) : t = 1, \dots, s\}, \quad Y(s, t) = \frac{1}{n_t} \sum_{r=1}^{n_t} X_r(t+r-1)$$

and  $n_t = \min(N, s-t+1)$ . The process  $Y$  consists of diagonal sums of variables in  $X$  as illustrated in Figure 9. This is the motivation for calling  $Y$  "diagonal process".

This transformation makes the surveillance problem a univariate problem! Observe however that in contrast to the situation usually considered in univariate surveillance, the variance of the statistic is not constant with respect to time. From time  $s$  to  $s+1$  the observations  $y(s, (s-N)^++1)$ ,  $y(s, (s-N)^++2)$ ,  $\dots$ ,  $y(s, s)$  are updated into  $y(s+1, (s-N)^++1)$ ,  $y(s+1, (s-N)^++2)$ ,  $\dots$ ,  $y(s+1, s)$  apart from the new observation  $y(s+1, s+1)$  at  $i=1$  being added. In other words, suppose  $s$  is larger than  $N$ , then having made observations  $y_{\leq s}$  the first  $s-N+1$  observations remain the same in  $y_{\leq s+1}$ , the last  $N-1$  in  $y_{\leq s}$  are updated by adding an  $x$  observation and normalising and finally also adding the observation  $x_1(s+1)$  made closest to the source.

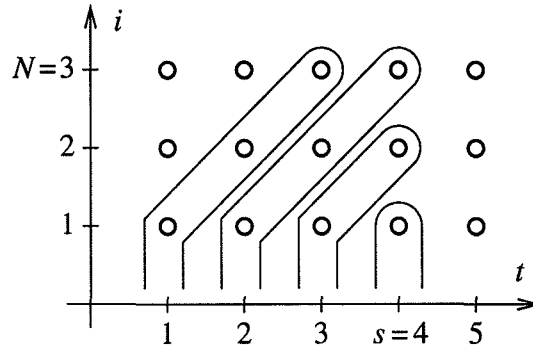


Figure 9: The diagonal process in the case when  $N=3$ . When  $s=4$  observations  $y(4, 1) = \min(4, 3)^{-1}(x_1(1) + x_2(2) + x_3(3))$ ,  $y(4, 2) = 3^{-1}(x_1(2) + x_2(3) + x_3(4))$ ,  $y(4, 3) = 2^{-1}(x_1(3) + x_2(4))$  and  $y(4, 4) = x_1(4)$  are made.

In the special case when  $N=1$  the diagonal process is the same as the observation process

$$Y_{\leq s} = \{X_1(1), \dots, X_1(s)\}$$

so this case corresponds to a well studied case of univariate surveillance.

Alternatively to constructing the diagonal process in the equidistant case as explained above, it may be convenient to use the following recursive construction of  $Y$ . For  $s \in \mathbb{Z}_+$  and  $t=1, \dots, s$

$$Y(s, t) = \begin{cases} X_1(1) & \text{if } s=1 \\ Y(s-1, t) & \text{if } s>1 \text{ and } N < s-t+1 \\ (s-t+1)^{-1}((s-t)Y(s-1, t) + X_{s-t+1}(s)) & \text{if } s>1 \text{ and } N \geq s-t+1. \end{cases}$$

This means that the last  $N-1$  variables in  $Y_{\leq s}$  change successively as  $s$  increases. Let us henceforth denote  $\{\mu(t) : t \leq s\}$  by  $\mu_{\leq s}$ .

**Remark 1**  $Y_{\leq s}$  is minimal sufficient for discrimination between  $C(s) = \{\tau \leq s\}$  and  $D(s) = \{\tau > s\}$ . This is a special case of Proposition 1 in Section 3.2.2.

From the construction of the diagonal process,  $Y$ , we have that conditional on  $\tau=t$ , all variables  $Y(s, r)$  in  $Y_{\leq s}$  are independent of each other and with  $n_r = \min(N, s-r+1)$

$$Y(s, r) \sim \begin{cases} N(0, n_r^{-1}) & \text{if } r < t \\ N(\delta, n_r^{-1}) & \text{if } r \geq t \end{cases}$$

The Shewhart method in this special case is

$$t_A = \min\{s \geq 1 : y(s, s) > c_1\}.$$



The Cusum method is

$$t_A = \min \left\{ s \geq 1 : \max_{1 \leq t \leq s} \sum_{r=t}^s n_r (y(s, r) - \delta/2) > c_2 \right\}.$$

Alternatively the Cusum method can be implemented as

$$t_A = \min \{ s \geq 1 : \max(v) > c_2 \}$$

where the vector  $v$  is defined recursively as

$$v(t) = \begin{cases} y(s, s) - \delta/2 & \text{if } t=1 \\ n_{s-t+1}(y(s, s-t+1) - \delta/2) + v(t-1) & \text{if } t=2, \dots, s. \end{cases}$$

The Shiryaev-Roberts method is

$$t_A = \min \left\{ s \geq 1 : \sum_{t=1}^s \exp \left( \delta \sum_{r=t}^s n_r (y(s, r) - \delta/2) \right) > c \right\}.$$

### Properties of the Methods Suggested for the Equidistance Case

To justify a comparison between the methods considered it is common to choose the threshold  $c$  for each method so that the expected time until false alarm,  $E[t_A | \tau > t_A]$  called  $ARL^0$ , is the same for all methods. For this reason we are interested in the relation between the threshold  $c$  and  $ARL^0$ .

The Shewhart threshold value is possible to calculate exactly for a fixed value of  $ARL^0$  as

$$c = \exp(\delta(\Phi^{-1}(1 - 1/ARL^0) - \delta/2)).$$

Throughout Section 3, all thresholds are chosen so that  $ARL^0 = 20$ . For the Cusum and Shiryaev-Roberts thresholds, simulations are used. The sample size is 1000 in these simulations such that the estimated values of  $ARL^0$  are within the interval (19.9, 20.1). Two features are revealed; first, for an unlimited number of sites in the lattice (i.e. " $N = \infty$ ") when  $\delta$  increases from 0.5 to 2, the threshold decreases (from 4.12 to 0.77 for the Cusum method and from 20.2 to 5.2 for the Shiryaev-Roberts method); second, when  $\delta$  is fixed and the number of sites (i.e.  $N$ ) increases, the thresholds decrease but convergence is very fast (in our simulations, thresholds for  $N = 10$  and " $N = \infty$ " were indistinguishable for shift sizes  $\delta = 0.5, 1, 1.5, 2$  respectively).

For comparing the surveillance methods one could look at the conditional stopping time distribution given that the change occurs at a time  $t$ . For the Shewhart method this is

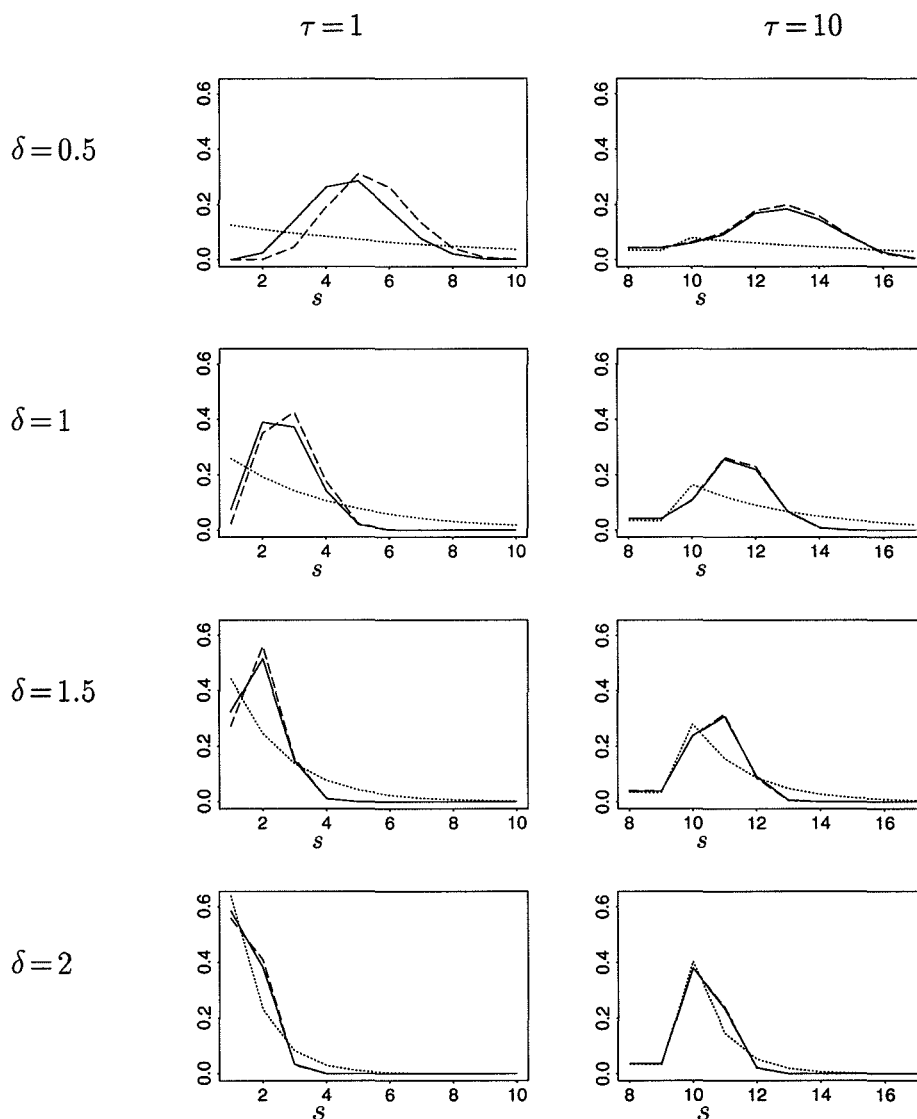


Figure 10: The conditional distribution,  $P[t_A = s | \tau = 1]$ , of  $t_A$  given that  $\tau = 1$  (left column) and  $P[t_A = s | \tau = 10]$ , given that  $\tau = 10$  (right column) for shift of size  $\delta = 0.5, 1, 1.5, 2$ . Dotted line indicates the Shewhart method, dashed line the Shiryaev-Roberts method and solid line the Cusum method. (Observe the truncations of the x-axes.)

$$\begin{aligned}
P[t_A = s | \tau = t] &= P[\{X_1(r) \leq c', r = 1, \dots, s-1\} \cap \{X_1(s) > c'\} | \tau = t] \\
&= \begin{cases} \Phi(c'')^{s-1}(1 - \Phi(c'')) & s < t \\ \Phi(c'')^{t-1}\Phi(c')^{s-t}(1 - \Phi(c')) & s \geq t \end{cases}
\end{aligned}$$

where  $c' = \delta^{-1} \log c + \delta/2$  and  $c'' = \delta^{-1} \log c - \delta/2$ . For the Cusum and the Shiryaev-Roberts methods these distributions were simulated with sample size 10 000. Plots of these distributions, when there is no limit,  $N$ , to the number of sites in the lattice (i.e. " $N = \infty$ "), can be seen in Figure 10.

A general impression is that the shapes of the performance measures considered are similar to the shapes of the corresponding measures in the case of a univariate variable shifting mean from 0 to  $\delta$  with constant variance (see e.g. Järpe and Wessman, 1999). When the shift size is larger than 1 all methods are quite similar. When the change occurs immediately, i.e. conditional on  $\tau = 1$ , the Shewhart method reacts quickest and when  $\delta = 0.5$  and 1, the Cusum method is slightly more likely to recognise this change sooner than the Shiryaev-Roberts method. In all other plots in Figure 10, the Cusum and the Shiryaev-Roberts methods are quite alike.

Another aspect of performance is the conditional expected delay of a motivated alarm, which is usually defined as  $E[t_A - \tau | t_A \geq \tau = t]$ .

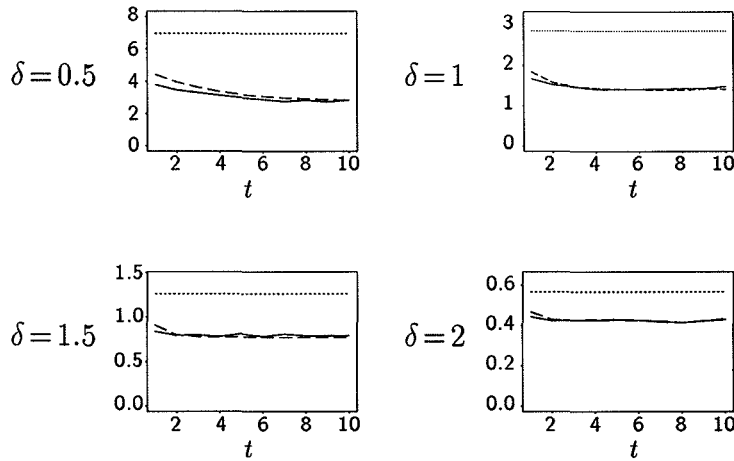


Figure 11: *The expected delay,  $E[t_A - \tau | t_A \geq \tau = t]$ , given that  $\tau = t$  for shift of size  $\delta$ . Dotted line indicates the Shewhart method, dashed line the Shiryaev-Roberts method and solid line the Cusum method. (Observe the different scales on the y-axes.)*

Since the Shewhart stopping time,  $t_A$ , is geometrically distributed with parameter  $1 - \Phi(\delta^{-1} \log c - \delta/2)$  when given that  $\tau = 1$ , the conditional expected

delay is

$$E[t_A - \tau | t_A \geq \tau = t] = (1 - \Phi(\delta^{-1} \log c - \delta/2))^{-1}$$

For the Cusum and Shiryaev-Roberts methods the expected delay was simulated (sample size 5 000) and all three methods plotted in Figure 11. From Figure 11 it is obvious that the Shewhart method is always the worst though the difference decreases as the shift size increases. The Cusum and Shiryaev-Roberts methods are similar except when the shift size is small,  $\delta = 0.5$  and when the change occurs immediately the Cusum method is slightly better than the Shiryaev-Roberts method.

## 4 Surveillance of Radiation

The consequences of a nuclear disaster can not be exaggerated. Rescue actions would have to be taken immediately after the incident and the precious time following afterwards would correspond directly to human lives. Therefore a reliable system for detecting a radiation change is needed.

### 4.1 The Problem

In 1986 there was a nuclear incident in Chernobyl, Russia, that was sensed far beyond the Russian borders. Later the same year, the Swedish radiation protection institute (SSI) completed the installation of 37 stations for measuring levels of gamma radiation. Figure 17, in Appendix C, shows data from five of these stations during 1998, each station making one observation every fifteen minutes. There was no shift in these data; the large increase in May in Övertorneå, Pajala and Kiruna is due to climate covariates. In May 26, at time 1.04.44 p.m., the value 405 nSv/h (nanosievert per hour) was reported from the Övertorneå station and in March 3, at time 12.50.43 a.m., the value 328 nSv/h was reported from the Kiruna station. These outliers have not been included in Figure 17 and will not be included in any plots henceforth. Nevertheless, they are included in the analysis of the data.

Fast and accurate surveillance methods are needed for making necessary preventive actions as soon as possible. For evaluation of these surveillance methods, some kind of restriction on the type of change is needed. One might consider an abrupt change in mean from one constant to another. Other shift types (such as a gradual shift spanning over some time interval) could be interesting and are discussed in Section 4.4.1. Here the former, a change from a constant mean  $\mu_0$  to another constant  $\mu_1 > \mu_0$ , is considered. Since no catastrophe happened in 1998 a shift was constructed artificially at time  $\tau = 0.00$  a.m. March 1 (see Figure 18 in Appendix C),  $\tau = 0.00$  a.m.

July 1, and  $\tau = 0.00$  a.m. November 1, for the purpose of illustrating different methods.

## 4.2 Covariates Explaining Variation of the Background Radiation

There are several covariates that might be interesting for explaining weather influence on data such as snow-depth, rain fall, temperature etc. An increase is discernible around April–May at stations 3, 4 and 5 (i.e. the plots in the second row of Figure 17) and a decrease during October–November at the same stations. These changes in radiation levels are annual and can partly be explained by the presence and absence of snow on the ground. When the radiation process is in control the stations measure background radiation i.e. mainly gamma particles that are radiated from the ground. Thus snow-depth is an interesting covariate since a layer of snow prevents particles from leaving the ground rendering lower radiation readings at the stations (see Figure 19, in Appendix C, for a picture of the snow-depths at the places of the radiation measuring stations).

The radiation data that was provided by SSI contained very few missing observations. However, the data on snow-depth contained were missing for large time periods. These data might thus not be totally reliable.

Because the snow accumulates during the winter season, it is mostly soft and porous during the first half year and increasingly hard and icy during the second. Therefore also a season dummy could be relevant together with snow-depth. This model could be described as

$$\Xi_i(t) = \beta_0 + \beta_1 S_i(t) + \beta_2 S_i(t) \times B(t) + \epsilon_i(t)$$

where  $S_i(t)$  is the snow-depth at site  $i$  and time  $t$ ,  $B(t)$  is 0 during the first half of the year and 1 during the second, and  $i = 1, 2, \dots, 5$ ,  $t = 1, 2, \dots$ . The residuals,  $x_i(t) = \xi_i(t) - \hat{\beta}_0 - \hat{\beta}_1 s_i(t) - \hat{\beta}_2 s_i(t) \times b(t)$ , from this regression are considered as observations of the spatial process.

As mentioned in Section 2, estimation of the covariate effects could be made during a “run-in” period previous to the actual surveillance. For this example, however, we have only observations from 1998 so, just for the purpose of illustration, the estimation is done from the same year (without the shift added) as the surveillance is performed (with the shift added) which of course would be impossible in real life.

A plot of the residuals with shift at March 1 added is in Figure 20 in Appendix C. Residuals are considered to be normally distributed in Section 3. Plots illustrating that this is not quite the case here are in Figure 12. This

might lead to non-valid theoretically derived properties of methods relying upon an assumption about normally distributed residuals. Thus it is important to have a model with many covariates that well reflect what can be explained by e.g. weather observations.

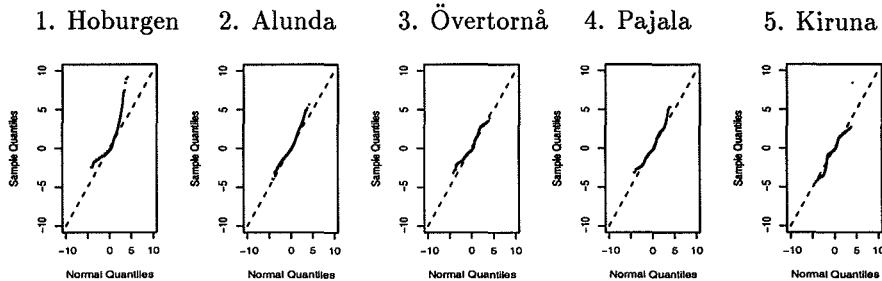


Figure 12: *Q-Q plots illustrating the deviance from the normal distribution of the residuals from regression with covariates.*

Other covariates that could be of interest are rainfall, topography, a function of time etc. For proper use, a more thorough investigation should be made in order to include as many relevant covariates as possible.

Having found all relevant covariates, the residuals from regression are assumed to be normally distributed, independent in all respects with global variances,  $\sigma^2$  (see Section 2). Having estimated the coefficients of this regression model, a shift of global size,  $\delta$ , in the observation process,  $\Xi$ , would remain the same size after having formed the spatial process,  $X$ , by subtracting the observed covariates according to the regression model. In this simple example however, we only consider snow-depth and a season dummy as covariate candidates. This means that there might still be dependence between variables at each site, deviance from normal distribution (such as skewness, heavy tails) and different variances for different sites.

After having cleaned the data from covariate influence, observations might be independent according to the model suggested in previous sections. To see to what extent this is fulfilled from the regression steps we calculate an estimate of the parameter,  $a$ , in an auto-regressive model for the residuals,  $X(t)$ , (i.e. residuals from the regression with covariates)

$$X(t) = aX(t-1) + \epsilon(t)$$

$t = 1, 2, \dots$  and  $\epsilon(t) \sim N(0, \sigma^2)$  independent in all respects. (Also an MA(1) process was tried but this had a worse fit than did the AR(1) model.) Plots of residuals, from first regression with covariates, shift at March 1 added, and then an AR(1) model, are in Figure 21 in Appendix C.

As can be seen from Table 1, the parameter estimates,  $\hat{a}$ , of  $a$  in the AR(1) model are quite close to 1 also after having cleaned from snow-depth and season, meaning either that there are important covariates that have not been included in the model, or that there is a strong genuine time dependence or both.

<i>Station</i>	Without covariates	With snow-depth		With snow-depth and season dummy		Percentage missing
	$\hat{a}$	$\hat{a}$	$R^2$	$\hat{a}$	$R^2$	
1. Hoburgen	0.873	0.871	0.017	0.870	0.028	0.17 %
2. Alunda	0.932	0.795	0.670	0.794	0.670	0.16 %
3. Övertorneå	0.979	0.859	0.852	0.857	0.854	0.13 %
4. Pajala	0.988	0.917	0.855	0.916	0.856	0.14 %
5. Kiruna	0.985	0.917	0.812	0.913	0.823	8.91 %

Table 1: *Estimates of the parameter,  $\hat{a}$ , in the AR(1) model before and after regression with snow-depth as a covariate, and determination coefficients,  $R^2$ , in the regression with snow-depth. Also the percentages of missing observations in the radiation data are given.*

To see to what extent the residuals, after regression with covariates and then an AR(1), model can be regarded as normally distributed, see the Q-Q plots in Figure 13. Several of the plots in Figure 13 indicate “heavy tail behaviour” compared to the normal distribution.

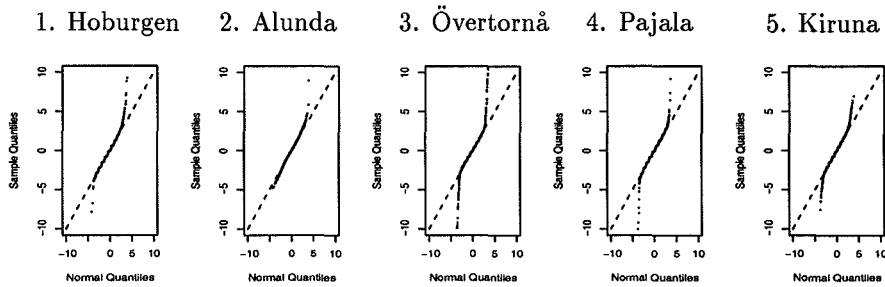


Figure 13: *Q-Q plots illustrating the deviance from the normal distribution of the residuals from first regression with covariates and then an AR(1) model.*

The residuals, from first regression with covariates, then with shift at March 1 added and finally an AR(1) model, are plotted in Figure 21 in Appendix C.

### 4.3 Radiation Shift Process

During 1998 there were no incidents and hence no shift to higher values due to this. In order to construct this shift, first a time of change,  $\tau$ , is fixed. Then, for each station  $i$ , a constant,  $\delta$ , is added to the observations made at or later than  $\tau+u_i$ . For adding the shift process, a spreading scenario must be considered: geography (i.e. mutual distances between sites and source), kind of spread (i.e. concentric, concentric with drift, line etc.). As an example we consider a shift originating from a source in Ignalina, Lithuania and spreading concentrically with a speed of 5 m/sec (see Figure 14). Here we consider three separate times of shift: either  $\tau$  is equal to time 0.00 a.m. March 1 (Figure 18 in Appendix C), or  $\tau$  is 0.00 a.m. July 1 or  $\tau$  is 0.00 a.m. November 1. The shift size,  $\delta$ , is three standard deviations of the unshifted spatial process. The standard deviations are calculated from residuals after having cleaned the observation process for covariates according to a linear model with snow-depth and a season dummy as covariates.

In the model in this paper, we consider shift in the mean of the spatial process,  $X$ , where the shift size,  $\delta$ , is the same for all sites,  $i \in A_N$ . We also let all processes be standardised, i.e.  $\text{Var}[X_i(t) - \mu_i(t)] = 1$  for all  $i \in A_N$ . A better modelling with several covariates might give about the same background variation. Thus for illustration the standardisation is made to artificially get the same variance, before and after the shift. This means that for the observation process,  $\Xi$ , the shift sizes for different sites may not be equal.

In the case of this example we have for the observation process that the shift size at Hoburgen is 8.44, at Alunda is 10.14, at Övertorneå is 15.86, at Pajala is 10.52 and at Kiruna the shift size is 17.01. A possible variation of the situation in this example for future studies could be to model the shift with a magnitude which is reversed proportional to the squared distance from the source.



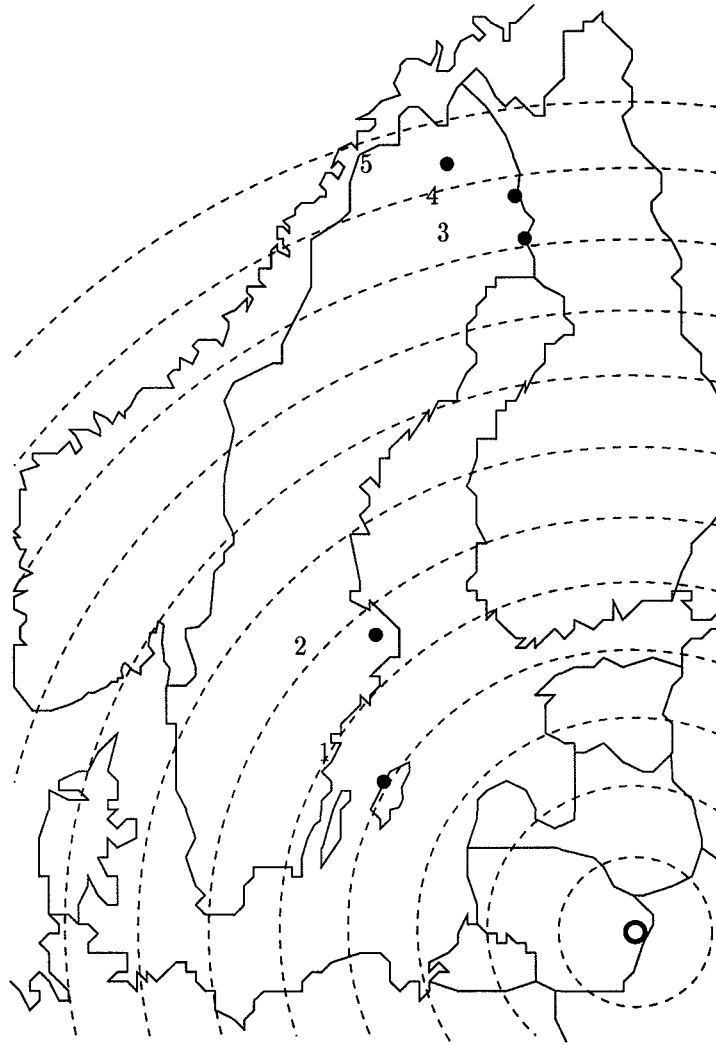


Figure 14: *The shift process. The distance between consecutive dashed arcs corresponds to 6 hours assuming that the shift spreads with 5 m/sec. If an incident should have its source in Ignalina this would mean a certain order for the shift process: first 1. Hoburgen, then 2. Alunda. After a 24 hour period 3. Övertorneå, 4. Pajala, and finally 5. Kiruna.*

#### 4.4 Methods for Surveillance of Radiation

Previously in this paper, we have been considering *active surveillance* (see e.g. Frisé and de Maré, 1991). Here, however, since an alarm is not expected to result in actions that will stop the increased radiation we have *passive surveillance*. In this situation we choose to compare observed times of delay of motivated alarms having calibrated the methods at a comparable rate of observed false alarms.

Some of the methods derived earlier in this report (see Section 3) are designed to handle spatial data and are thus dealing with all stations simultaneously. This is the case for methods based upon the diagonal process. Other methods are designed to handle just one station. For these “single station methods” one approach is to use the union-intersection principle and make an alarm as soon as any one of the stations signals alarm. Denoting the local alarm function for station  $i$  by  $p_i(s)$ , using the union-intersection principle renders an alarm function,  $p(s) = \max_i p_i(s)$  for each time  $s = 1, 2, 3, \dots$ . As an alternative, a very simple method which takes the spatial structure into account is one which only considers values from the station closest to the source. Denoting the local alarm function for station  $i$  by  $p_i(s)$ , using this “nearest location principle” renders an alarm function,  $p(s) = p_1(s)$  for each time  $s = 1, 2, 3, \dots$  (if station 1 is closest to the source). As mentioned in Sections 3.2 and 3.2.2, this coincides with one of the definitions of the Shewhart method for the diagonal process.

The basic methods considered here are the *window method*, the *Shewhart method*, and the *Cusum method*. First we describe the presently used method, the *window method* in Section 4.4.1. Then, we illustrate the likelihood ratio methods of the earlier sections.

In the earlier sections time dependence was not treated. For surveillance of the mean radiation levels, time dependence may be handled in different ways. Here we consider two approaches, one which is called *likelihood ratio method with threshold adjustment for an auto-regressive model* (in Section 4.4.2), and another called *likelihood ratio method applied on residuals from an auto-regressive model* (in Section 4.4.3). Thus the methods under consideration can be further categorised into the following 12 schemes.

1. **Win NL**: window method applied to the single stations and then nearest location principle for getting a global method,
2. **Win UI**: window method applied to the single stations and then union-intersection principle for getting a global method,
3. **Thresh NL**: Shewhart method applied residuals from regression with

covariates for the single stations and then nearest location principle for getting a global method (this is the same as the Shewhart method for the diagonal process based on residuals from regression),

4. **Thresh UI**: Shewhart method applied residuals from regression with covariates from the single stations and then union-intersection principle for getting a global method,
5. **Thresh ARNL**: Shewhart method applied residuals from first regression with covariates and an auto-regressive model for the single stations and then nearest location principle for getting a global method (this is the Shewhart method for the diagonal process based on residuals from first regression and then an auto-regressive model),
6. **Thresh ARUI**: Shewhart method applied residuals from first regression with covariates then an auto-regressive model from the single stations and then union-intersection principle for getting a global method,
7. **Cus NL**: Cusum method applied residuals from regression with covariates from the single stations and then nearest location principle for getting a global method,
8. **Cus UI**: Cusum method applied residuals from regression with covariates from the single stations and then union-intersection principle for getting a global method,
9. **Cus ARNL**: Cusum method applied residuals from first regression with covariates and an auto-regressive model for the single stations and then nearest location principle for getting a global method,
10. **Cus ARUI**: Cusum method applied residuals from first regression with covariates then an auto-regressive model for the single stations and then union-intersection principle for getting a global method,
11. **Cus diag**: Cusum method applied the diagonal process based on residuals from regression,
12. **Cus ARdiag**: Cusum method applied the diagonal process based on residuals from first regression and then an auto-regressive model.

These are the categories considered in the comparison in Section 4.4.4 where the performance of different methods for the surveillance of the radiation data is considered. The Cusum method applied to these data is not fully evaluated and the Shiryaev-Roberts method is not included in Section 4

(except for the pseudo-code given in the Appendix) since calculations of their alarm functions are quite heavy.

To justify the comparison, thresholds are chosen so that the observed false alarm rates for each station, do not exceed that of the window method at the corresponding stations. Then the observed delays of motivated alarms are compared. Also, the thresholds were, in some cases, chosen higher so that the false alarm rate was much lower than that of the window method at the corresponding station, when this did not increase observed delays of motivated alarm.

#### 4.4.1 Window Method

SSI today uses an alarm rule (Kjelle, 1987), henceforth called the *window method* which makes an alarm as soon as the difference between the sums of data from two consecutive 24 hours periods exceeds a prescribed threshold,  $L$ . Kjelle illustrated the performance of the window method by some examples using 100, 200 and 300 as values of  $L$ . (However, no theoretical properties of this method were reported in that paper.) In the US Food and Drug Administration guidelines (1991), a window method (the difference between the present observation and the mean of previous observations during a fixed time period, as alarm function) was suggested. Disadvantages of that method was demonstrated by Sveréus (1995). Other uses of moving window statistics were studied by Roberts (1966), Stroup et al (1989), Wallenstein and Neff (1987), and in a spatial setting by Kulldorff (1997). Figure 22, in Appendix C, shows a plot of the window method alarm function applied directly to the radiation data with shift at March 1. Results of using this window method with the threshold at level 100 are given in Table 2. The delay times from the local shifts are all less than 24 hours, most of them less than 10 hours but none of them less than 3 hours. This implies first motivated alarm times more than 35 hours from when the shift occurred.

Shift	Station	False alarms/month	Delay from local shift	Delay from source shift
March 1	Hoburgen	53	15 h.	40 h.
	Alunda	82.5	8 h. 15 min.	41 h. 30 min.
	Övertorneå	27	5 h. 30 min.	63 h. 15 min.
	Pajala	14	7 h. 15 min.	69 h. 15 min.
	Kiruna	43	9 h. 15 min.	88 h. 15 min.
July 1	Hoburgen	70	10 h. 30 min.	35 h. 30 min.
	Alunda	27.5	8 h. 30 min.	41 h. 15 min.
	Övertorneå	22	6 h.	64 h. 15 min.
	Pajala	23.3	9 h.	71 h.
	Kiruna	72.5	5 h. 30 min.	71 h. 30 min.
Nov. 1	Hoburgen	67.4	11 h. 15 min.	36 h. 15 min.
	Alunda	33.2	9 h. 15 min.	43 h.
	Övertorneå	41.6	4 h.	62 h. 15 min.
	Pajala	54.8	5 h.	67 h.
	Kiruna	77.4	3 h.	69 h.

Table 2: *False alarms/month and delay of motivated alarm for local window methods.*

The approach used today by SSI does not systematically take into account covariates like snow depth etc. The other methods for surveillance in Section 4 are all applied to residuals from regression and residuals from first regression and then an auto-regressive model as explained in Section 4.2. The reason for not applying the window method to residuals from such regressions but rather to the raw radiation data directly is that this is what is done today for the monitoring of data deriving from these stations. Therefore all comparisons must be seen in the light of this. An obvious disadvantage with

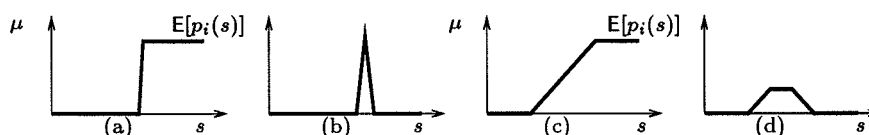


Figure 15: *With the window method a sudden shift of a process (graph (a)) would be transformed into a peak of the alarm function,  $p$ , (graph (b)). However, if the shift was gradual (graph (c)), the window method would transform this into a less sudden but smaller shift which then shifts back (graph (d)).*

the window method is that once the window has passed the time of shift, if

we have a sudden change from one constant level to another constant, there is no indication of high radiation levels any more. For this kind of change the window method always results in a few high values when the first window is mostly before the change and the other is mostly after the change (see Figure 15). A disadvantage with any method based directly upon data without cleaning from covariate effects, is that one cannot tell whether high values are due to a beginning of a catastrophe or are due to a change in covariates e.g. a decrease in snow-depth.

This window method does not provide a systematic procedure for weighing together information from different stations. Thus we use the union-intersection principle and the "nearest location principle" mentioned in the beginning of Section 4.4. The results of applying the union-intersection principle to the window method (indicated by *Win UI*) are given in Table 7 and applying the "nearest location principle" (indicated by *Win NL*) in Table 8 in Section 4.4.4. Even though based on raw radiation data, the window method is not always the worst. When compared to methods based upon residuals from a better regression model, however, the situation might be quite different.

#### 4.4.2 Likelihood Ratio Methods With Empirical Threshold Adjustment for an Auto-regressive Model

One way of dealing with possible auto-regressive dependencies without forming residuals from an auto-regressive model is the following.

For the choice of thresholds one would like as few false alarms and as little delay of a motivated alarm as possible. For the sake of a comparison between methods, the thresholds were chosen so that the observed number of false alarms would not be greater than the corresponding number for the window methods with threshold 100.

The Shewhart alarm function for each station when the shift occurs at March 1, is given by the plot of the residuals and threshold values in Figure 20 in Appendix C. Observed delay of motivated alarm may be read from Table 3. The delay times from the local shifts are very small. In most cases there is no delay of alarm at this rate of false alarms.

Shift	Station	False alarms/month	Delay from local shift	Delay from source shift
March 1	Hoburgen	53	1 h. 15 min.	26 h. 15 min.
	Alunda	82	none	33 h. 15 min.
	Övertorneå	26.5	none	58 h. 15 min.
	Pajala	12.5	15 min.	62 h. 15 min.
	Kiruna	17	none	66 h.
July 1	Hoburgen	62	none	25 h.
	Alunda	26.8	15 min.	33 h. 30 min.
	Övertorneå	20.2	none	58 h. 15 min.
	Pajala	20.7	none	62 h.
	Kiruna	43	none	66 h.
Nov. 1	Hoburgen	62	none	25 h.
	Alunda	31.2	1 h.	34 h. 15 min.
	Övertorneå	31.3	none	62 h.
	Pajala	40.8	none	62 h.
	Kiruna	77.4	none	66 h.

Table 3: *False alarms/month and delay of motivated alarm for local Shewhart methods (with an empirical threshold adjustment) for the three times of shift.*

Values of the “nearest location principle” applied to these singular station alarm function values are given in Table 7 (labelled *Thresh NL*) and the union-intersection principle applied to the Shewhart in Table 8 (labelled *Thresh UI*). The low delay times are preserved when using the “nearest location principle” (which makes the method identical to the Shewhart method for the diagonal process) and the union-intersection principle.

As can be seen from Figures 14 and 16, the spatial structure is far from equidistant. In fact from 263 intervals there are 5 intervals with one site in each and 258 without any sites.

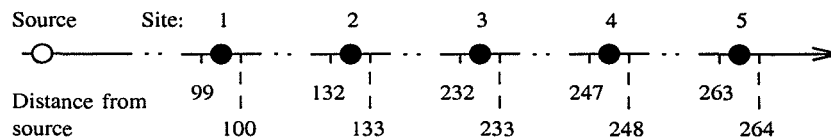


Figure 16: The spatial structure of this example is far from equidistant.

For the Shewhart method this means no problems but for the the Cusum and the Shiryaev-Roberts methods it has an effect on the properties. The

definition of the diagonal process in Section 3.2.2

$$Y(s, t) = n_t^{-1} \sum_{r=1}^{n'_t} \sum_{i:u_i=r} X_i(t+r-1),$$

where  $n_t = \#\{i : u_i \leq s-t+1\}$  and  $n'_t = \min(\max_i u_i, s-t+1) - \min_i u_i + 1$ , is used. However, calculation of the alarm function values for both the Cusum and the Shiryaev-Roberts methods for the diagonal process was very heavy. These values were therefore not plotted together with thresholds as were other methods. There are two exceptions though: one is the Cusum alarm function for the diagonal process based upon the residuals from regression with shift at March 1, plotted in Figure 24 in Appendix C, and values for comparison with other methods (indicated by *Cus diag*) are in Tables 7 and 8 in Section 4.4.4. The other is the Cusum alarm function for the diagonal process based upon residuals, from first regression and then an AR(1) model, which is discussed further in Section 4.4.3. Pseudo-code for implementing these methods in some low level programming language is given in Appendix B.

Another way of using the marginal data from each station would be to apply Cusum locally. The Cusum alarm function for each station on its own with shift at March 1, with thresholds adjusted so that the false alarm rates do not exceed the corresponding false alarm rates of the window method with threshold 100, is plotted in Figure 23 in Appendix C. The false alarm rates and delay times are given in Table 4. The Cusum method is more sensitive to small but systematic changes before time  $s$  than the Shewhart method. For corresponding rates of false alarm, the delay times are longer (see Figure 12). The reason for the long delay of alarm compared with the Shewhart method is possibly that the regression model needs improvements. Due to poor adjustment for covariates, there is not a constant mean.

Values of the "nearest location principle" applied to these singular station alarm function values are given in Table 7 (labelled *Cus NL*) and the union-intersection principle applied to the Cusum in Table 8 (labelled *Cus UI*).



Shift	Station	False alarms/month	Delay from local shift	Delay from source shift
March 1	Hoburgen	52.5	19 h. 30 min.	44 h. 30 min.
	Alunda	75.5	7 h.	40 h. 15 min.
	Övertorneå	0	none	58 h. 15 min.
	Pajala	13	11 h. 45 min.	73 h. 45 min.
	Kiruna	39	30 min.	66 h. 30 min.
July 1	Hoburgen	69.5	15 h.	40 h.
	Alunda	25.2	13 h. 15 min.	46 h. 30 min.
	Övertorneå	22	> 10 days	> 10 days
	Pajala	23.3	14 h. 30 min.	76 h. 30 min.
	Kiruna	35.7	none	66 h.
Nov. 1	Hoburgen	67.4	12 h. 45 min.	37 h. 45 min.
	Alunda	33.2	11 h. 45 min.	45 h.
	Övertorneå	41.6	> 9 days	> 9 days
	Pajala	53.1	12 h. 15 min.	74 h. 15 min.
	Kiruna	77.4	51 h. 45 min.	117 h. 45 min.

Table 4: *False alarms/month and delay of motivated alarm for local Cusum methods (with an empirical threshold adjustment) for the three times of shift.*

In the residuals from regression there is an increase in radiation in Övertorneå during May and in Kiruna during October not explained by snow-depth data. These are due to the fact that data on snow-depth do not correspond so well to the radiation data during these two periods at the respective stations according to the linear model suggested. This in turn may be because of two omitted readings of snow-depth. If these increases in the residuals had been the beginning of harmful shifts to higher levels, they would have been sensed by the Cusum method for the diagonal process. It illustrates the sensitivity of the Cusum method and the importance of having accurate covariates in the regression.

One very conservative way of dealing with AR(1) dependence would be to modify the stopping rule thresholds and use the same Shewhart procedure but with threshold  $c' = c(1 - a^2)^{-1/2}$ . This would guarantee a higher  $ARL^0$  than in the independent case with threshold  $c$  (see e.g. Pettersson, 1998). However, when the auto-regressive model parameter,  $a$ , is close to 1, this method is not so useful since it tends to put the thresholds too far away for the delays of motivated alarms to be reasonable.

#### 4.4.3 Likelihood Ratio Methods Applied on Residuals from an Auto-regressive Model

To get independent observations one could form successive residuals. However, this would also transform a shift of the mean (see Pettersson, 1998) which is a drawback for surveillance purposes.

Consider an AR(1) model for residuals from regression

$$X(t) = aX(t-1) + \epsilon(t).$$

In Figure 21 in Appendix C is a plot of the residuals,  $\hat{\epsilon}_i(t)$ , from first regression and then adding shift at March 1 and finally residuals from an auto-regressive model together with threshold values corresponding to the same or fewer observed false alarms per month as with the window method with threshold 100, indicated by dashed horizontal lines. In Table 5 there are some values of observed delay of motivated alarm of the Shewhart method applied to these residuals. Since most of the shift is deleted by the forming of

Shift	Station	False alarms/month	Delay from local shift	Delay from source shift
March 1	Hoburgen	50	2 h. 15 min.	27 h. 15 min.
	Alunda	80	15 min.	33 h. 30 min.
	Övertorneå	26	8 h. 15 min.	66 h. 30 min.
	Pajala	12	189 h. 45 min.	251 h. 45 min.
	Kiruna	30.5	4 h. 45 min.	70 h. 45 min.
July 1	Hoburgen	63.7	2 h.	27 h.
	Alunda	27.3	194 h.	227 h. 15 min.
	Övertorneå	20.8	11 h. 30 min.	69 h. 45 min.
	Pajala	21.3	3 h. 30 min.	65 h. 30 min.
	Kiruna	55.5	none	66 h.
Nov. 1	Hoburgen	67.2	2 h.	27 h.
	Alunda	32.9	73 h.	106 h. 15 min.
	Övertorneå	35.5	11 h. 45 min.	70 h.
	Pajala	52.7	30 min.	62 h. 30 min.
	Kiruna	33.3	none	66 h.

Table 5: *False alarms/month and delay of motivated alarm for local Shewhart methods (applied to residuals from first regression and then an auto-regressive model) for the three times of shift.*

AR(1) residuals, the delay times are higher than compared to the case with

residuals from only regression (see Table 3). The forming of residuals from an auto-regressive model changes both the variance and the shift sizes. For making the methods comparable, these residuals were standardised so that the variance was approximately unity.

Values of the “nearest location principle” applied to the Shewhart single station alarm function values are given in Table 7 (labelled *Thresh ARNL*) and the union-intersection principle applied to the Shewhart single station method in Table 8 (labelled *Thresh ARUI*). The forming of AR(1) residuals increases the delay times for the Shewhart method in this case since most of the shift is lost.

Also here one may apply a local Cusum on the marginal residuals from first regression and then an auto-regressive model. The Cusum alarm functions for each station on its own where the shift occurs at March 1, with thresholds such that the false alarm rates do not exceed the corresponding false alarm rates of the window method with threshold 100, are plotted in Figure 25 in Appendix C. The false alarm rates and delay times are given in Table 6.

Shift	Station	False alarms/month	Delay from local shift	Delay from source shift
March 1	Hoburgen	50.5	15 h. 45 min.	40 h. 45 min.
	Alunda	79.5	3 h. 15 min.	36 h. 30 min.
	Övertorneå	26	none	58 h. 15 min.
	Pajala	14	8 h. 15 min.	70 h. 15 min.
	Kiruna	43	none	66 h.
July 1	Hoburgen	68.8	7 h. 15 min.	32 h. 15 min.
	Alunda	27.5	7 h.	40 h. 15 min.
	Övertorneå	21.7	> 10 days	> 10 days
	Pajala	22.8	2 h. 30 min.	64 h. 30 min.
	Kiruna	2.2	none	66 h.
Nov. 1	Hoburgen	67.2	9 h. 15 min.	34 h. 15 min.
	Alunda	33.2	27 h. 45 min.	61 h.
	Övertorneå	41.4	> 9 days	> 9 days
	Pajala	53.5	7 h.	59 h.
	Kiruna	77.4	17 h.	83 h.

Table 6: *False alarms/month and delay of motivated alarm for local Cusum methods (applied to residuals from an auto-regressive model) for the three times of shift.*

Even though the forming of AR(1) residuals is removing most of the shift, there is a small but systematic shift left. In contrast to the Shewhart method, the Cusum method takes observations from the past into account and the delay times are therefore improved compared to the Cusum method applied to residuals from only regression. Nevertheless, the unity variance is very large compared to the size of the shifts so the delay times are high compared to the Shewhart method.

Due to the increase in the residuals from the regression, mentioned in Section 4.4.2, there is a systematic deviance from 0 in the residuals, from first regression and then the AR(1) model (this is barely visible in Figure 21 in Appendix C). So again the sensitivity of the Cusum method and the importance of having accurate covariates in the regression, are illustrated.

Values of the “nearest location principle” applied to the Cusum single station alarm function values are given in Table 7 (labelled *Cus ARNL*) and the union-intersection principle applied to the Cusum in Table 8 (labelled *Cus ARUI*). The performance of the Cusum method is slightly improved by forming AR(1) residuals but unfortunately to the cost of most of the shift.

The Cusum method for the diagonal process based upon the residuals from first regression, then adding a shift in March 1, and finally an auto-regressive model, is plotted in Figure 26 and values of observed delay of alarm are in Tables 7 and 8 in Section 4.4.4 (labelled *Cus ARdiag*). Also for the diagonal process of the AR(1) residuals, the performance of the Cusum method is deteriorated by small shifts compared to high variance. There is also a systematic deviance during May at the Övertorneå station (discussed more in Section 5) which also affects the performance of the Cusum method.

#### 4.4.4 Evaluation Summary

The Shewhart method applied residuals from regression for the single stations and then applied “nearest location principle” (called *thresh NL*) is superior in all situations considered. The *window methods* (both from applying the “nearest location principle” and the union-intersection principle) are inferior in many of the situations considered.

The Cusum method applied to the diagonal process, in the non-equidistant case was explained in Section 3.2.2. For shift in March 1, Figures 24 and 26, in the Appendix, show plots of the Cusum alarm function (where the diagonal processes use residuals from regression with covariates and residuals from first regression and then an auto-regressive model). For the case when the shift occurs in March 1, false alarm rates and delays of these two methods for the diagonal process are in Tables 7 and 8. The Cusum method for the diagonal process formed by residuals from first regression with covariates and

Shift	Method	False alarms/month	Delay from local shift	Delay from source shift
March 1	Win NL	53	15 h.	40 h.
	Thresh NL	53	1 h. 15 min.	26 h. 15 min.
	Thresh ARNL	50	2 h. 15 min.	27 h. 15 min.
	Cus NL	52.5	19 h. 30 min.	44 h. 30 min.
	Cus ARNL	50.5	15 h. 45 min.	40 h. 45 min.
	Cus diag	52.5	19 h. 45 min.	44 h. 45 min.
	Cus ARdiag	53	17 h. 30 min.	42 h. 30 min.
July 1	Win NL	70	10 h. 30 min.	35 h. 30 min.
	Thresh NL	62	none	25 h.
	Thresh ARNL	63.7	2 h.	27 h.
	Cus NL	69.5	15 h.	40 h.
	Cus ARNL	68.8	7 h. 15 min.	32 h. 15 min.
Nov. 1	Win NL	67.4	11 h. 15 min.	36 h. 15 min.
	Thresh NL	62	none	25 h.
	Thresh ARNL	67.2	2 h.	27 h.
	Cus NL	67.4	12 h. 45 min.	37 h. 45 min.
	Cus ARNL	67.2	9 h. 15 min.	34 h. 15 min.

Table 7: *False alarm rates and delay times for the nearest location methods considered in Section 4 and, for shift at March 1, Cusum for the diagonal process.*

Shift	Method	False alarms/month	Delay from local shift	Delay from source shift
March 1	Win UI	219.5	15 h.	40 h.
	Thresh UI	161	1 h. 15 min.	26 h. 15 min.
	Thresh ARUI	198.5	2 h. 15 min.	27 h. 15 min.
	Cus UI	180	11 h. 15 min.	36 h. 15 min.
	Cus ARUI	215	11 h. 30 min.	36 h. 30 min.
	Cus diag	140	15 min.	25 h. 15 min.
	Cus ARdiag	146.5	none	25 h.
July 1	Win UI	215.3	10 h. 30 min.	35 h. 30 min.
	Thresh UI	173	none	25 h.
	Thresh ARUI	188.6	2 h.	27 h.
	Cus UI	175.7	15 h.	40 h.
	Cus ARUI	143	7 h. 15 min.	32 h. 15 min.
Nov. 1	Win UI	274.4	11 h. 15 min.	36 h. 15 min.
	Thresh UI	201	2 h.	27 h.
	Thresh ARUI	221.6	2 h.	27 h.
	Cus UI	272.7	12 h. 45 min.	37 h. 45 min.
	Cus ARUI	272.7	9 h. 15 min.	34 h. 15 min.

Table 8: *False alarm rates and delay times for the union-intersection methods considered in Section 4 and, for shift at March 1, Cusum for the diagonal process.*

then an auto-regressive model, was specified for shift of size 0.5.

## 5 Discussion

Optimality of methods for surveillance of measures on several sites depends on the kind of catastrophes to detect and on the dependency structure. The solutions for some such combinations will now be discussed.

The case when optimal properties are desired for detection of a shift in mean at all (or a known subset of) sites at the same time (or with a known time-lag) can, by using earlier results, be reduced to surveillance of a univariate statistic. This is as expected since there is only one random change-point. Given this, the times of shift for all processes are given. Ordinary methods of surveillance such as the Cusum or the Shewhart methods can thus be directly applied to the derived statistic.

If nothing is known about the time-lag between shifts at different locations, then we have a general multivariate surveillance. No uniformly optimal method is available but several good methods have been suggested in the literature and they are applied for this situation. The dependency structure at each time-point is especially important in this case.

The case of a local change, at one or a few sites, with a later spread to include an ever larger cluster of sites, is considered. The time of the first change is random but, given this, the time-points of the other changes are supposed to be known. A statistic which is minimal sufficient for the surveillance is derived. With this statistic the spatial surveillance problem is reduced to a univariate one. Shewhart, Cusum and Shiryaev-Roberts surveillance methods for this statistic are derived and briefly evaluated by means of simulations. Since the statistic does not have constant variance with respect to time, the properties of the above methods are not the same as those generally described. Having chosen the expected time until false alarm as 20, the evaluation is made with respect to the distribution of the time of alarm given that the first change occurs immediately and given that the first change occurs at time 10. Also, the expected delay, given that the change occurs at times  $1, 2, \dots, 10$ , is given. The shapes of these curves look similar to the corresponding ones for the common univariate case, with independence between observations given the time of change. The most prominent messages from the plots are that the Cusum and Shiryaev-Roberts methods are superior to the Shewhart method with respect to expected delay, and that the Cusum and Shiryaev-Roberts methods are quite similar.

An example of application is surveillance of radiation. In 1998, a collaboration between the Department of Statistics, Göteborg University and the

Swedish Radiation Protection Institute was initiated. Data from five stations measuring gamma radiation in Sweden, are used to give illustrations of how the methods perform. The possibilities of utilising covariates are illustrated by the use of a very simple regression model. A more subject matter knowledge based modelling could be expected to increase the efficiency of the surveillance. Comparisons are made for each location between the window method used today, the Shewhart method and the Cusum method. Notable about the window method is that once the window has passed the time of shift, if we have a sudden change from one constant level to another constant, there is no indication of high radiation levels any more. The window method is based upon data which have not been cleaned from covariate effects. A disadvantage with any method based directly upon raw data is that one cannot tell whether high values are due to a beginning of a catastrophe or are due to a change in covariates e.g. a decrease in snow-depth.

For simultaneous surveillance of all locations one general method (not utilising any information about time-lag between sites), one simple method (utilising all information about the nearest location) and the new suggested diagonal method (utilising all information about the time-lags) are considered. From the evaluation, the Shewhart method for the diagonal process seems preferable to the window method used today. The Shewhart method for the diagonal process had the shortest delay from local shift as well as delay from the source shift at a comparable rate of false alarms for all three times of shift and among all methods considered. However, this cannot be regarded as a comparison between the window method and the Shewhart method but rather a comparison between the method used today and some alternatives (since the evaluation of the window method is made based upon raw data of radiation levels, while the evaluation of all other methods is based on residuals from regression). Considering that the window method is based upon raw data rather than residuals from a regression, lots of difficulties with regression are avoided. Nevertheless, with a proper regression model conditions for surveillance using the Shewhart, Cusum or Shiryaev-Roberts methods would be expected to improve radically. Still, with the very simple regression model used in Section 4, the window method is inferior in most cases, in respect of observed delay of alarm from local shift as well as from source shift.

Surprisingly, the Cusum method (both applied to local data and to the diagonal process) seems surprisingly not superior to the Shewhart method with respect to observed delay of alarm from local shift or from source shift. The Cusum method for the diagonal process is slightly better than the compared union-intersection methods considered but not better than the compared nearest location methods. In the Övertorneå residuals from regression there



is a monotone increase during the beginning of May and in the Kiruna residuals from regression there is a monotone increase during October. This results in a small but systematic deviance from 0 in residuals from first regression and then an auto-regressive model. This deviance is due to available snow data not corresponding so well to the radiation data according to the linear model suggested. This in turn may be because of a few omitted readings of snow-depth. If these small increases in the residuals had been the beginning of a harmful shift to higher levels, it would have been sensed by the Cusum method for the diagonal process. Then, very likely, the Cusum method for the diagonal process would be superior to the Shewhart method in the sense of having a shorter (expected) delay for a fixed (expected) number of false alarms as indicated by the results in Section 3.2.2. This may serve as an illustration of the sensitivity of the Cusum method and the importance of having accurate covariates in the regression.

The comparisons are intended just as illustrations. For evaluation of properties one replicate would not be enough but simulation studies or analytically derived results are necessary.

Further investigations on how to treat a process which possesses spatial as well as temporal dependencies would be useful. Examples are studies of the situation where the shift size decays with increasing distance to the source and use of other general surveillance methods.

## 6 Appendix

### 6.1 Appendix A: Proofs

**Proposition 1**  $Y_{\leq s}$  is minimal sufficient for discrimination between  $C(s) = \{\tau \leq s\}$  and  $D(s) = \{\tau > s\}$ .

**Proof:** Conditional on  $\tau = t$ , the likelihood ratio,  $L(s, t)$ , of  $x_{\leq s}$  given  $\tau = t \leq s$  versus given  $\tau > s$  is minimal sufficient for

$$\mu_{\leq s} = \{\delta I(t \leq r - u_i) : r = 1, \dots, s \text{ and } i \in A_N\}$$

and therefore

$$\log L(s, t) = \log \frac{f(x_{\leq s} | \tau = t)}{f(x_{\leq s} | \tau > s)} = \delta \sum_{r=t}^s \sum_{i: u_i \leq r-t+1} (x_i(r) - \delta/2)$$

is minimal sufficient as well. Hence the likelihood ratio

$$\frac{f(x_{\leq s} | \tau \leq s)}{f(x_{\leq s} | \tau > s)} = \exp \left( \sum_{t=1}^s \pi_t L(s, t) \right)$$

is sufficient for  $\mu_{\leq s}$  unconditional on  $\tau$  (where  $\pi_t = P[\tau = t]$ ). We may write the log likelihood triangle  $\log L(s, t)$  (or polygon if  $\max_i u_i > s - t$ ) as a sum of vertical slices where each slice is a sum of  $x_i(r)$  components. But this triangle is also possible to slice diagonally and thus express in terms of  $y(s, r)$  since

$$\begin{aligned}
\log L(s, t) &= \delta \sum_{r=t}^s \sum_{i: u_i \leq r-t+1} (x_i(r) - \delta/2) \\
&= \delta \left( \sum_{i: u_i \leq 1} (x_i(t) - \delta/2) + \sum_{i: u_i \leq 2} (x_i(t+1) - \delta/2) + \dots \right. \\
&\quad \left. + \sum_{i: u_i \leq s-t+1} (x_i(s) - \delta/2) \right) \\
&= \delta \left( \sum_{r=1}^{s-t+1} \sum_{i: u_i = r} (x_i(t+r-1) - \delta/2) + \dots \right. \\
&\quad \left. + \sum_{r=1}^1 \sum_{i: u_i = r} (x_i(t+r-1) - \delta/2) \right) \\
&= \delta (n_t(y(s, t) - \delta/2) + n_{t+1}(y(s, t+1) - \delta/2) + \dots + n_s(y(s, s) - \delta/2)) \\
&= \delta \sum_{r=t}^s n_r(y(s, r) - \delta/2)
\end{aligned}$$

which is a linear combination of

$$\{y(s, t), y(s, t+1), \dots, y(s, s)\}.$$

Thus we have that the full likelihood ratio

$$\sum_{t=1}^s \pi_t L(s, t) = \sum_{t=1}^s \pi_t \exp \left( \delta \sum_{r=t}^s n_r (y(s, r) - \delta/2) \right)$$

is a one-to-one function of

$$y_{\leq s} = \{y(s, t), y(s, t+1), \dots, y(s, s) : t = 1, 2, \dots, s\},$$

and therefore  $Y_{\leq s}$  is minimal sufficient for discrimination between  $C(s)$  and  $D(s)$ .

□

**Proposition 2** *In a lattice structure of  $B_m$ ,*

$$Y(s, t) \sim \begin{cases} N(0, v_t) & t < \tau \\ N(\delta, v_t) & t \geq \tau \end{cases}$$

where  $(\min(N, s-t+1)+m)^{-1} \leq v_t \leq (\min(N, s-t+1)-m')^{-1}$  and  $m' = \min(m, N, s-t+1)$ .

**Proof:** Let  $B_m$  be the family of lattice structures which differ from the equidistant case by exactly  $m$  sites (see Section 3.2.2). Due to the construction of the diagonal process,  $E[Y(s, t)|\tau] = \delta I(t \geq \tau)$ . For the simple structure  $B_0$  with  $N$  sites, one site in each unit interval, the variables at time  $t$  are

$$X_1(t), X_2(t), \dots, X_N(t).$$

Suppose that  $m$  sites are changed (at  $m$  occasions either a variable is added or removed) rendering the variables at time  $t$

$$X'_1(t), X'_2(t), \dots, X'_{N'}(t)$$

where  $N-m \leq N' \leq N+m$  ( $N' = N-m$  if a variable was removed at all occasions, and  $N' = N+m$  if a variable was added at all occasions). Then for the diagonal process variables,  $Y(s, t)$ , formed from this lattice structure

$$\begin{aligned} \text{Var}[Y(s, t)] &= n_t^{-2} \sum_{r=1}^{n'_t} \sum_{i:u_i=r} \text{Var}[X_i(t+r-1)] \\ &= n_t^{-2} (m_1 \text{Var}[X_1(t)] + m_2 \text{Var}[X_2(t+1)] + \dots \\ &\quad + m_{n'_t} \text{Var}[X_{n'_t}(t+n'_t-1)]) \end{aligned}$$

where  $m_i = \#\{j : u_j = i\}$ . But  $\text{Var}[X_i(t)] = 1$  for all  $i$  and  $t$ ,

$$n'_t - m' \leq m_1 + m_2 + \dots + m_{n'_t} \leq n'_t + m$$

and  $n_t = m_1 + m_2 + \dots + m_{n'_t}$ . Thus we have that

$$(n'_t + m)^{-1} \leq \text{Var}[Y(s, t)] \leq (n'_t - m')^{-1}.$$

The statistic  $Y(s, t)$  is normally distributed since it is a sum of normal random variables.

□

## 6.2 Appendix B: Program Code

For applying the Cusum method using the diagonal process in the situation of the radiation example in Section 4, the following pseudocode might be useful.

```

maxdist    ← {number of time units between
              site closest to source and site farthest}
alarmfcn   ← {0 : i = 1, ..., tmax}
alarmfcn(1) ← (y(1,1) - δ/2)+
temp       ← {0 : i = 1, ..., tmax}
for s = 2 to tmax
  diago     ← {0 : i = 1, ..., s}
  lo        ← max(1, 1 - maxdist + 1)
  for t = lo to s
    diago(t) ← n(t,s) · (y(t,s) - δ/2)
  temp(1)   ← alarmfcn(s-1) + diago(1)
  for t = lo to s
    temp(t-lo+2) ← sum({diago(i) : i = t - lo + 1, ..., s})
  alarmfcn(s) ← max(0, temp)
return alarmfcn

```

The pseudocode for the Shiryaev-Roberts method applied to the diagonal process is the following.

```

maxdist    ← {number of time units between
              site closest to source and site farthest}
alarmfcn   ← {0 : i = 1, ..., tmax}
alarmfcn(1) ← lr(1,1)
temp       ← {0 : i = 1, ..., tmax}
hist       ← = 0
now        ← = 0
for s = 2 to tmax
  if s > maxdist
    now ← (hist + 1) · lr(s - maxdist, s)
  next   ← 1
  lo     ← max(1, 1 - maxdist + 1)
  for t = lo to s
    next ← next · lr(s - t + 1, s)

```

```
    alarmfcn(s) → now + next
    hist       → now
return alarmfcn
```

where  $lr(t, s)$  is a function which returns  $\exp(\delta n(t, s)(y(t, s) - \delta/2))$ .

### 6.3 Appendix C: Figures

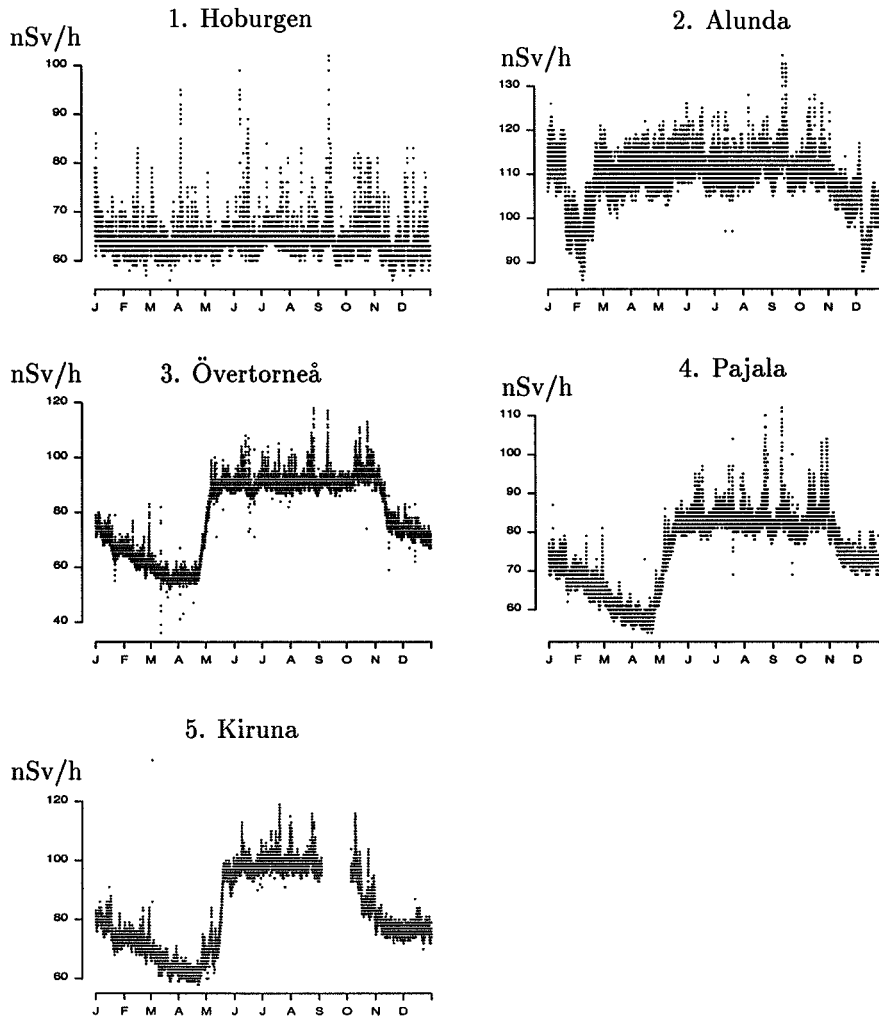


Figure 17: Radiation data during 1998 from 5 of the 37 stations located in Sweden.

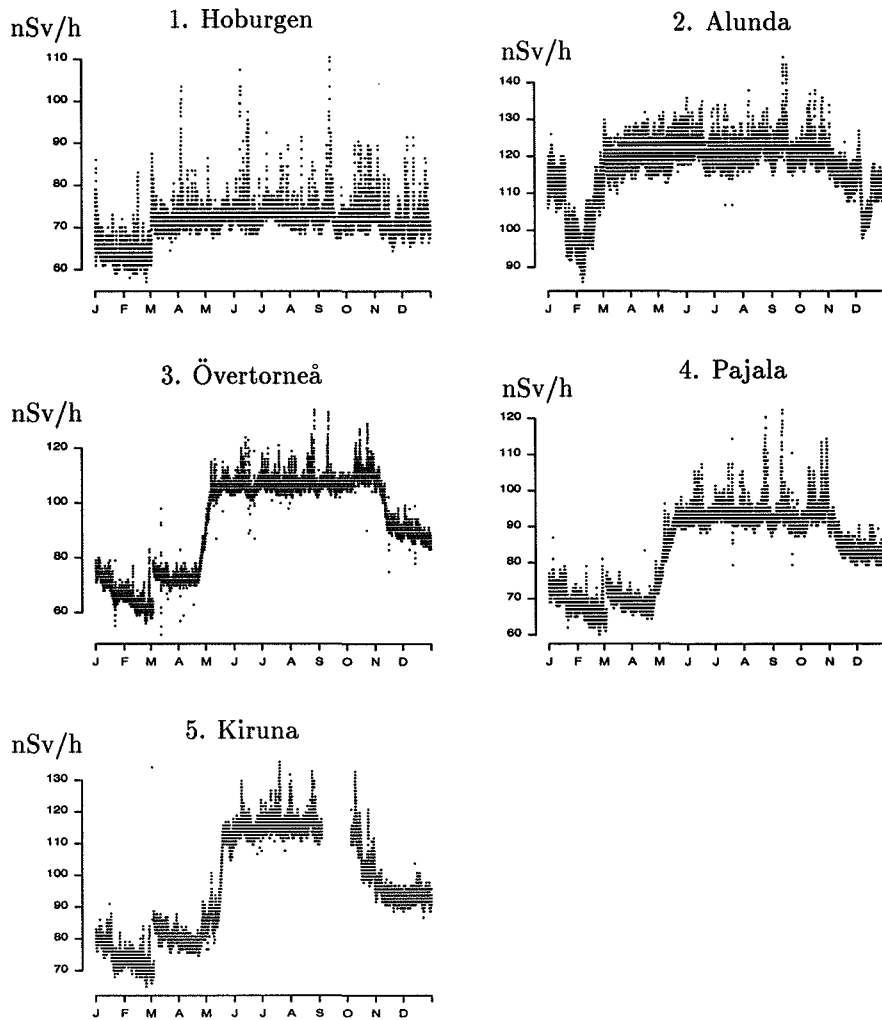


Figure 18: *The same data as in Figure 17 but with a shift in mean of the size of three standard deviations originating in March 1 added.*

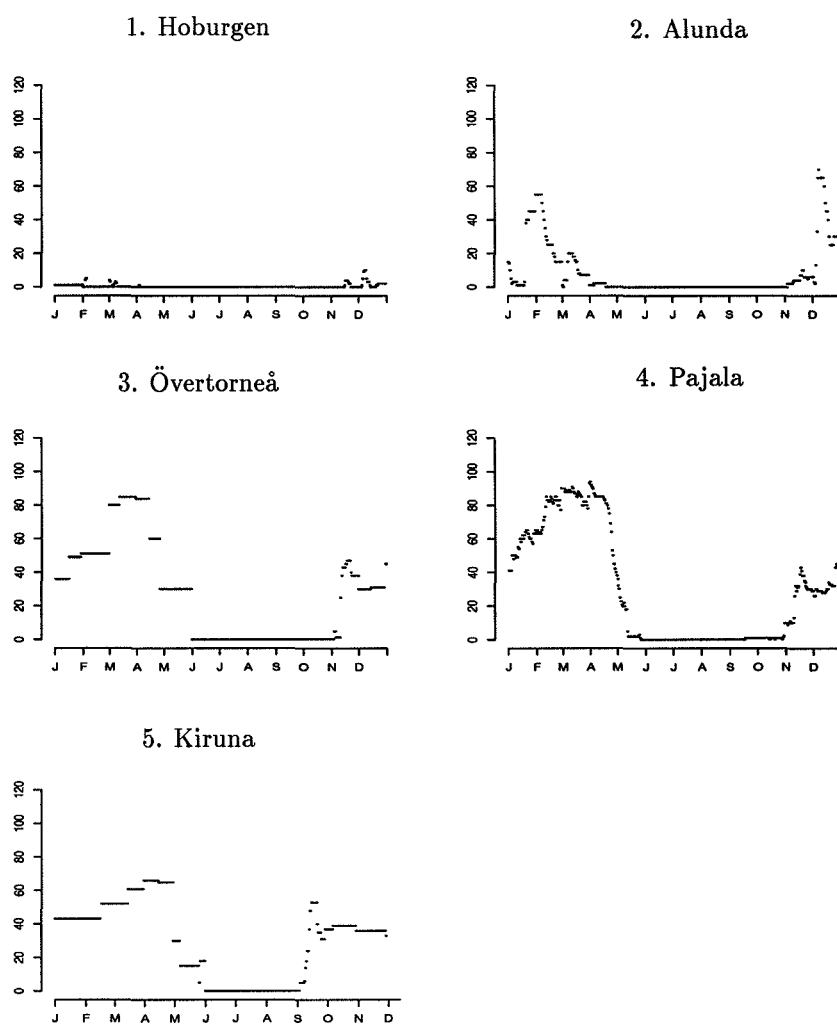


Figure 19: Snow-depth data during 1998.



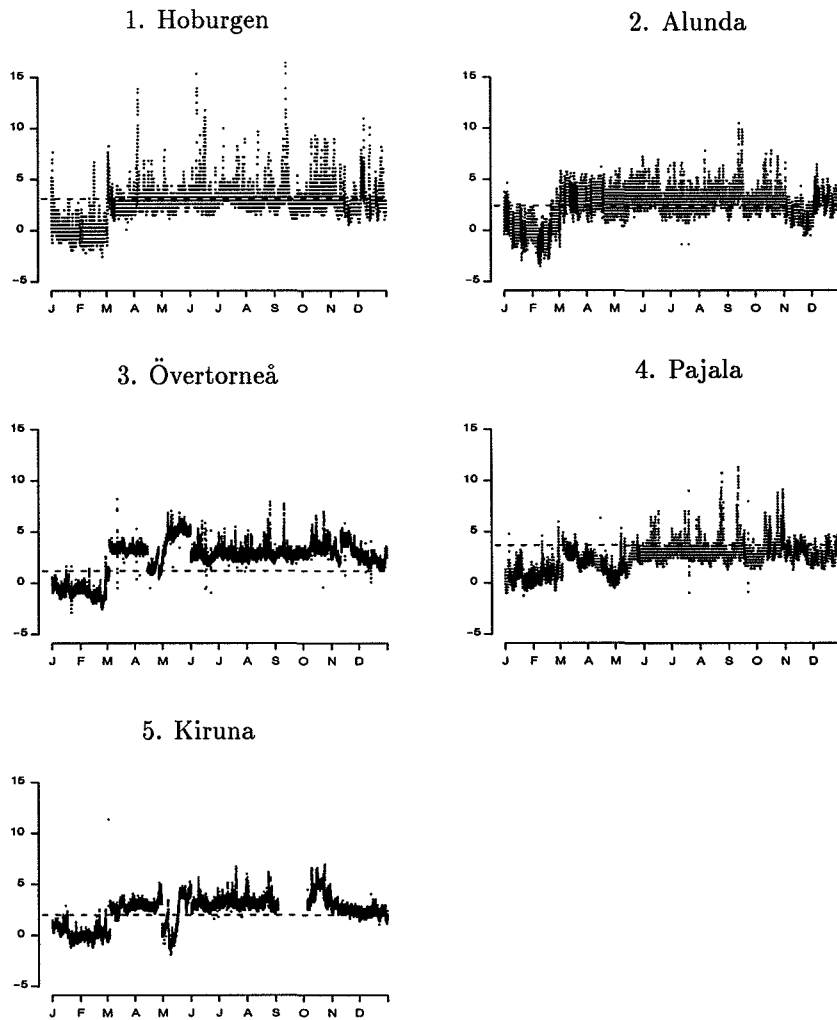


Figure 20: *The residuals from regression according to the model with snow-depth and snow-depth  $\times$  season as covariates. The time of shift is March 1. The dashed lines indicate thresholds giving fewer false alarms for the local Shewhart method than does the window method with thresholds 100 at each corresponding station.*

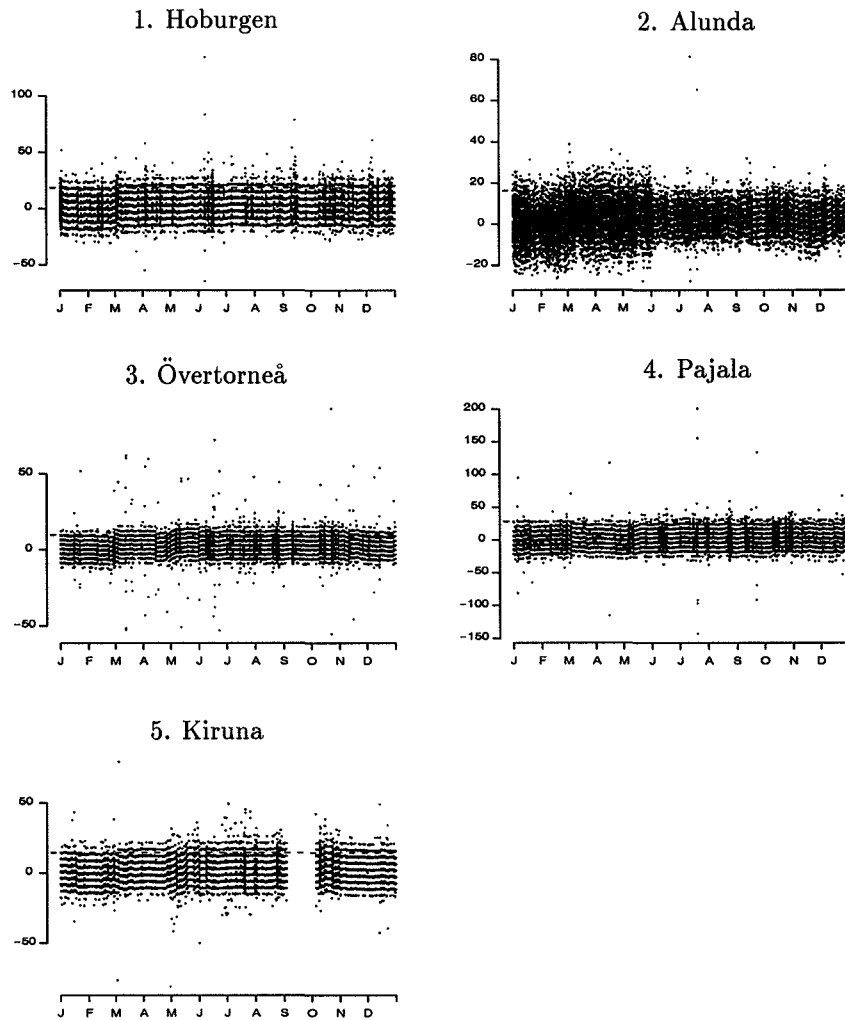


Figure 21: *The residuals from first regression with snow-depth as a covariate and then an auto-regressive process of the five respective stations. The shift is at March 1. The dashed lines indicate thresholds giving fewer false alarms for the local Shewhart method than does the window method with thresholds 100 at each corresponding station.*

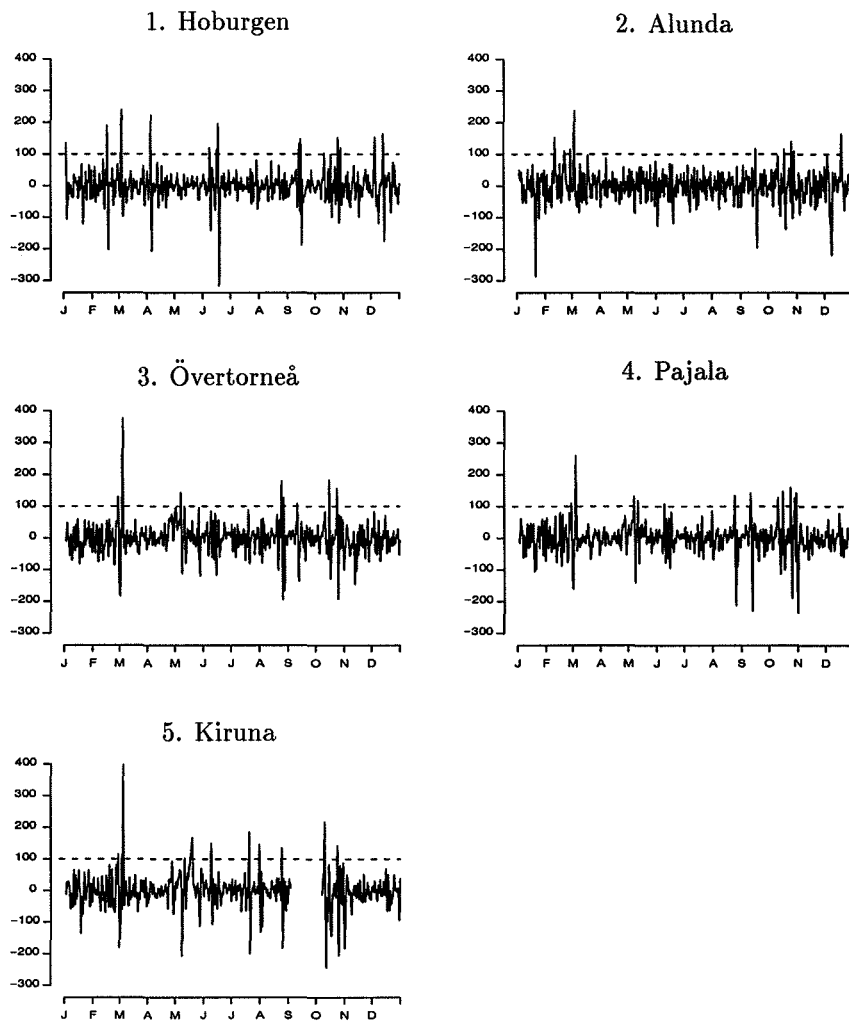


Figure 22: Values of the SSI alarm function illustrating the window method in this example. The dashed line indicates a threshold value at level 100. The shift is at March 1.

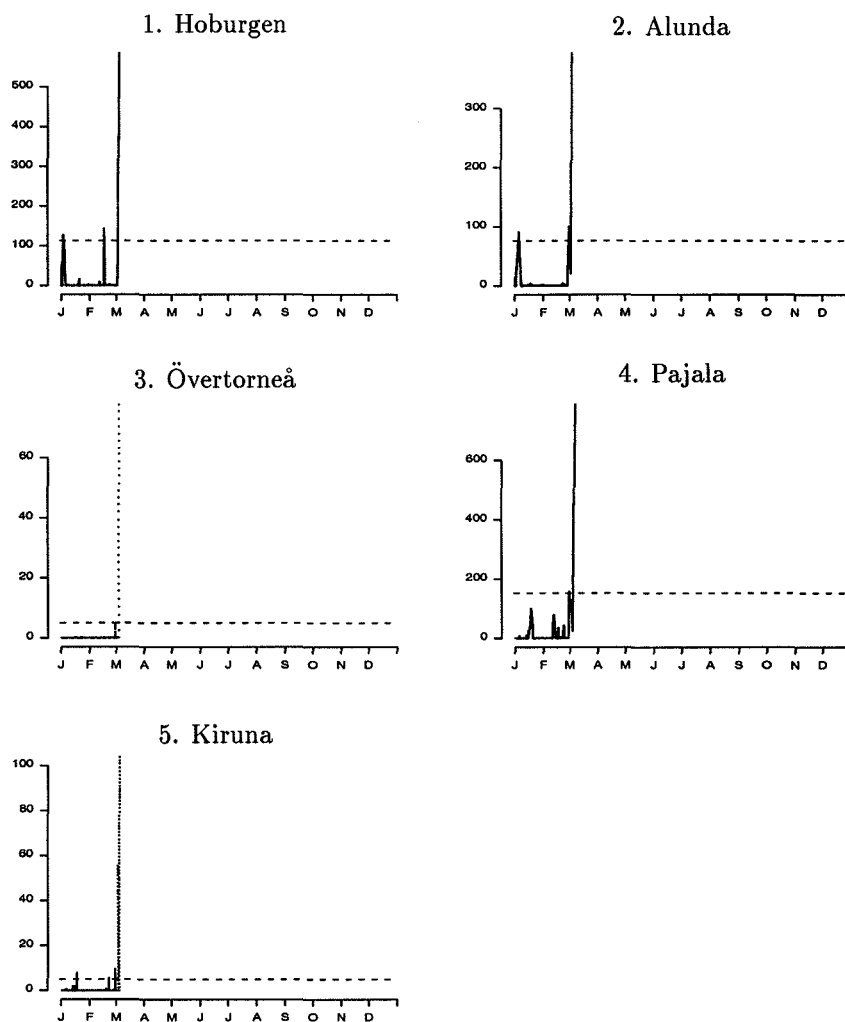


Figure 23: *The Values of the Cusum alarm function applied to residuals from regression with snow-depth and season as covariates. The shift is at March 1. The dashed lines indicate thresholds giving fewer false alarms for the local Cusum method than does the window method with thresholds 100 at each corresponding station.*

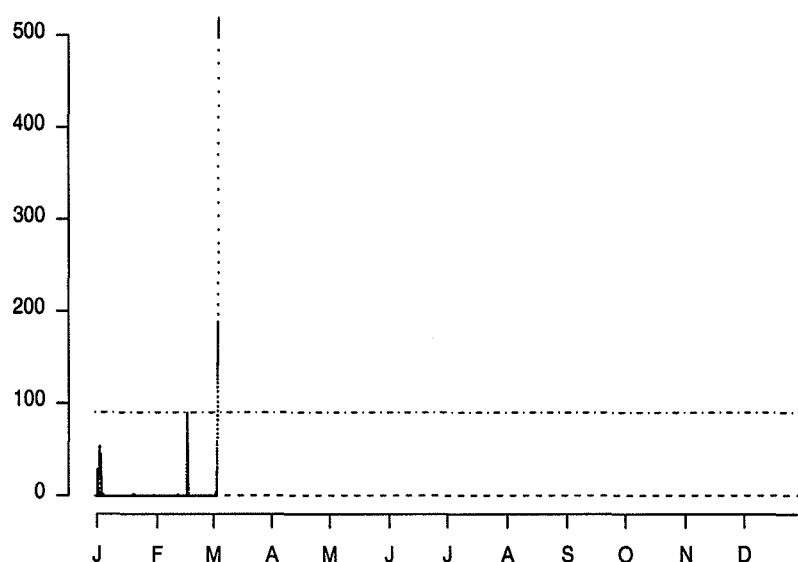


Figure 24: *The values of the Cusum alarm function for the diagonal process constructed of residuals from regression with covariates (referred to as Cus diag in Tables 7 and 8). The shift is at March 1. The dashed line indicates a threshold giving fewer false alarms for the Cusum method than does the union-intersection principle applied to the window method with thresholds 100. The dash-dotted line indicates a threshold giving fewer false alarms for the Cusum method than does the “nearest location principle” applied to the window method with thresholds 100.*

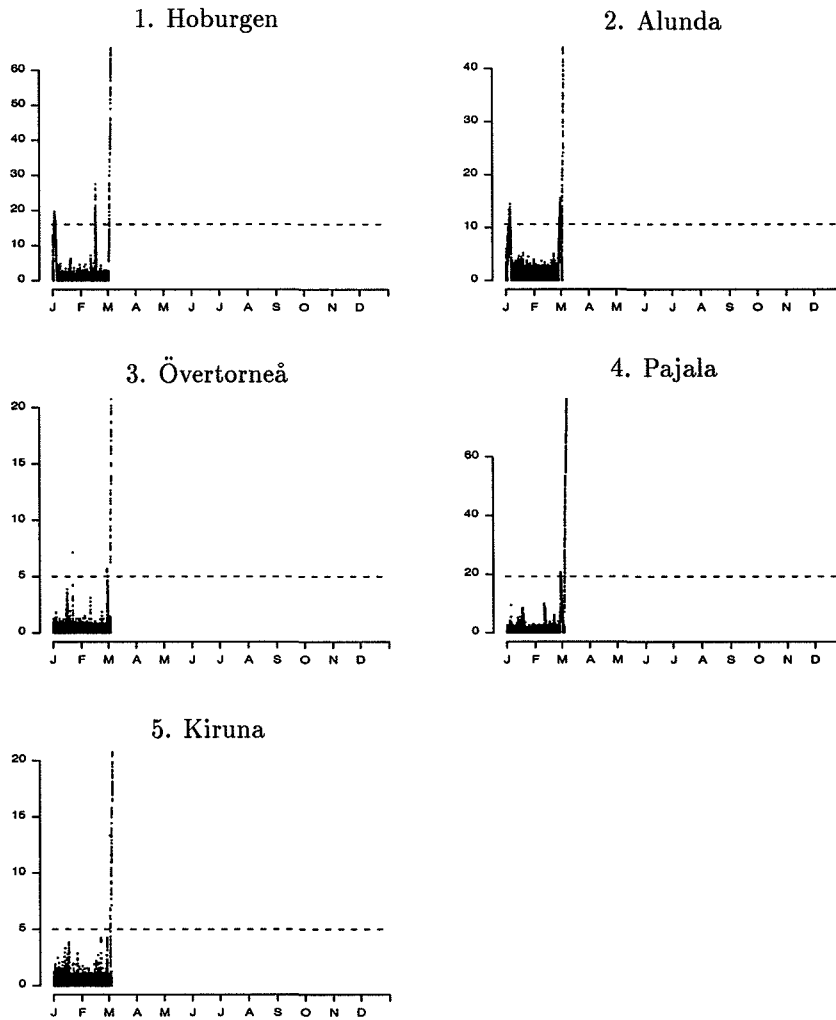


Figure 25: *The Values of the Cusum alarm function applied to residuals from first regression with snow-depth and season as covariates and then an auto-regressive model. The shift is at March 1. The dashed lines indicate thresholds giving fewer false alarms for the local Cusum method than does the window method with thresholds 100 at each corresponding station.*

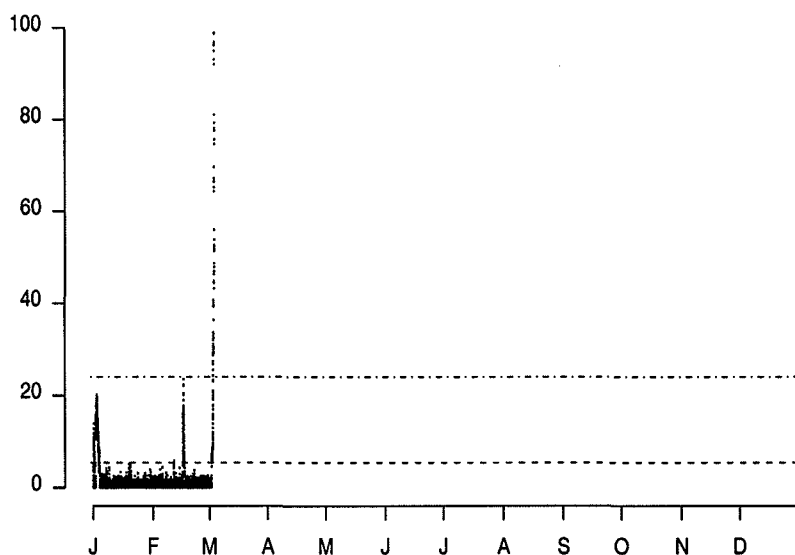


Figure 26: *The values of the Cusum alarm function for the diagonal process constructed of residuals from first regression with covariates and then an auto-regressive model (referred to as Cus ARdiag in Tables 7 and 8). The shift is at March 1. The dashed line indicates a threshold giving fewer false alarms for the Cusum method than does the union-intersection principle applied to the window method with thresholds 100. The dash-dotted line indicates a threshold giving fewer false alarms for the Cusum method than does the “nearest location principle” applied to the window method with thresholds 100.*

## Acknowledgements

I am very grateful to Professor Marianne Friséen for reading several previous versions of this paper and for offering brilliant suggestions. I would also like to thank Associate Professor Aila Särkkä for a great deal of insightful advice and patient readings of my manuscript efforts and Christian Sonesson for interesting and fruitful discussions. This work was supported by the Swedish Radiation Protection Institute and the Swedish Council for Research in the Humanities and Social Sciences.

## REFERENCES

- Cressie, N.A.C. (1993) *Statistics for Spatial Data*, Wiley, USA.
- Draft Guideline for Post Marketing Reporting of Adverse Drug Reactions (1991) *US Food and Drug Administration*, Rockville, MD.
- Friséen, M. (1999) Characterisation of Surveillance by Optimality, *Research Report 1999:13*, Göteborg University, Göteborg.
- Friséen, M. and de Maré, J. (1991) Optimal Surveillance, *Biometrika*, vol. 78, no. 2.
- Friséen, M. and Wessman, P. (1999) Evaluations of Likelihood Ratio Methods for Surveillance: Differences and Robustness, *Communications in Statistics, Simulation and Computation*, vol. 28, no. 3.
- Järpe, E. (2000) Surveillance of Spatio-Temporal Patterns: Change of Interaction in an Ising Dynamic Model, *Research Report 2000:7*, Göteborg University, Göteborg.
- Järpe, E. (1999) Surveillance of the Interaction Parameter of the Ising Model, *Communications in Statistics, Theory and Methods*, vol. 28, no. 12.
- Järpe, E. and Wessman, P. (1999) Some Power Aspects of Methods for Detecting Different Shifts in the Mean, *Communications in Statistics, Computation and Simulation*, vol. 29, no. 2.
- Kjelle, P.E. (1987) Alarm Criteria for the Fixed Gamma Radiation Monitoring Stations, *SSI report 87-07*, Swedish Radiation Protection Institute.
- Kulldorff, M. (1997) A Spatial Scan Statistic, *Communications in Statistics, Theory and Methods*, vol. 26., no. 6.
- Lai, T.L. (1995) Sequential Change-Point Detection in Quality Control and Dynamical Systems. *Journal of the Royal Statistical Society*, vol. 57.
- Lawson, A., Biggeri, A., Böhning, D., Lesaffre, E., Viel, J.-F., and Berollioni, R. (1999) *Disease Mapping and Risk Assessment for Public Health*, Wiley, Chichester.



- Moustakides, G.V. (1986) Optimal Stopping Times for Detecting Changes in Distributions, *Annals of Statistics*, vol. 14, no. 4.
- Page, E.S. (1954) Continuous Inspection Schemes, *Biometrika*, vol. 41, pp 100–115.
- Pettersson, M. (1998) Evaluation of Some Methods for Statistical Surveillance of an Autoregressive Process, *Research Report 1998:4*, Göteborg University, Göteborg.
- Roberts, S.W. (1966) A Comparison of Some Control Chart Procedures, *Technometrics*, vol. 8, no. 3.
- Rogerson, P.A. (1997) Surveillance Systems for Monitoring the development of Spatial Patterns, *Statistics in Medicine*, vol. 16.
- Roy, S.N. (1953) On a Heuristic Method of Test Construction and Its Use in Multivariate Analysis, *Annals of Mathematical Statistics*, vol. 24.
- Shewhart, W.A. (1931) Economic Control of Quality of Manufactured Product, *Van Nostrand*, New York.
- Shiryayev, A.N. (1963) On Optimum Methods in Quickest Detection Problems, *Theory of Probability and its Applications*, vol. 8.
- Siegmund, D. (1985) Sequential Analysis. Tests and Confidence Intervals, *Springer-Verlag*, New York.
- Srivastava, M. and Wu, Y. (1997) Evaluation of Optimum Weights and Average Run Lengths in EWMA Control Schemes, *Communications in Statistics, Theory and Methods*, vol. 26, no. 5.
- Srivastava, M. and Wu, Y. (1993) Comparison of EWMA, CUSUM, and Shiryaev-Roberts Procedures for Detecting a Shift in the Mean, *Annals of Statistics*, vol. 21, no. 2.
- Stroup, D.F. et al (1989) Detection of Aberrations in the Occurrence of Notifiable Diseases Surveillance Data, *Statistics in Medicine*, vol. 8.
- Sveréus, A. (1995) Detection of Gradual Changes. Statistical Methods in Post Marketing Surveillance, *Research Report 1995:2*, Göteborg University, Göteborg.
- Wallenstein, S. and Neff, N. (1987) An approximation for the Distribution of the Scan Statistic, *Statistics in Medicine*, vol. 6.
- Wessman, P. (1998) Some Principles for Surveillance Adopted for multivariate Processes with a Common Change Point, *Communications in Statistics, Theory and Methods*, vol. 27.
- Wessman, P. (1999) The Surveillance of Several Processes with Different Change Points, *Research Report 1999:2*, Göteborg University, Göteborg.
- Whittle, P. (1954) On Stationary Processes in the Plane, *Biometrika*, vol. 41.
- Åkermo, G. (1994) On Performance of Methods for Statistical Surveillance, *Research Report 1994:7*, Göteborg University, Göteborg.

## Research Report

- |         |  |   |
|---------|--|---|
| 1999:5  | Mantalos, P. & Shukur, G.:             | Testing for cointegrating relations- A bootstrap approach.  |
| 1999:6  | Shukur, G.:                            | The effect of non-normal error terms on the properties of systemwise RESET test.  |
| 1999:7  | Järpe, E. & Wessman, P.:               | Some power aspects of methods for detecting different shifts in the mean.   |
| 1999:8  | Johnsson, T.:                          | On statistics and scientific thinking.  |
| 1999:9  | Afsarinejad, K.:                       | Trend-free repeated measurement designs.  |
| 1999:10 | Carlquist, A. m.fl.                    | The impact of season and climate on growth during early childhood in different socio-economic groups in Lahore, Pakistan. |
| 1999:11 | Carlquist, A, Erling, V. & Frisén, M.: | Longitudinal methods for analysis of the influence of breastfeeding on early child in Pakistan.                           |
| 1999:12 | Carlquist, A.:                         | Longitudinal methods for analysis of early child health in Pakistan.  |
| 1999:13 | Frisén, M.:                            | Characterization of methods for surveillance by optimality.   |
| 2000:1  | Hatemi-J, A. & Shukur, G.:             | Multivariate based causality tests of twin deficits in the US.  |
| 2000:2  | Petzold, M.:                           | Aspects on tolerance limit estimation - some common approaches and flexible modeling.                                     |
| 2000:3  | Andersson, L.:                         | Statistical test of the existence of a turning point.   |
| 2000:4  | Almasri, A.:                           | Some aspects of wavelet analysis in time series.  |
| 2000:5  | Järpe, E.                              | Surveillance of spatio-temporal patterns. Change of interaction in an Ising dynamic model.                                |