Daumer, Bauer, Lederer:

Average Run Length and Mean Delay for Changepoint Detection: Robust Estimates for Threshold Alarms

Sonderforschungsbereich 386, Paper 95 (1997)

Average Run Length and Mean Delay for Changepoint Detection: Robust Estimates for Threshold Alarms

M. Daumer†, C. Bauer‡, C. Lederer§

14th October 1997

† Institute for Medical Statistics, Technical University of Munich 81675 Munich, Germany
‡ Institute for Statistics, Technical University of Munich, Germany
§ Institute for Mathematics, Humbold University, Berlin, Germany

1 Introduction

Online Monitoring is a rapidly expanding field in different areas such as quality control, finance and navigation. The automated detection of so-called changepoints is playing a prominent role in all these fields, be it the detection of sudden shifts of the mean of a continuously monitored quantity, the variance of stock quotes or the change of some characteristic features indicating the malfunctioning of one of the detectors used for navigation (the "faulty sensor problem").
A prominent example for the application of advanced statistical methods for the detection of changepoints in biomedical time series is the multi-process Kalman filter used by Smith and West [Smith 1983] to monitor renal transplants. However, despite the fact that the algorithm could be tuned in such a way that the computer could predict dangerous situations on the average one day before the human experts it has nevertheless become superfluous as soon as new diagnsoic tools became available.
Many of the automated monitoring systems which are widely used in practice are based on simple threshold alarms. Some upper and lower limits are chosen at the beginning of the monitoring session and an alarm is triggered whenever the measured values exceed the upper limit or fall below the lower limit. This is e.g. common practice for the monitoring of patients during surgery, where such thresholds are chosen for heart rate, blood pressure etc. by the anaesthesist.
The fate of the multi-process Kalman filter for monitoring renal transplants teaches two lessons: first, there is considerable power in statistical methods to

improve conventional biomedical monitoring techniques. Second, if the statistical model and the methods are too refined they may never be used in practice.
We shall suggest a stochastic model for changepoints which we have found to have the capacity to be very useful in practice, i.e. which is sufficiently complex to cover the important features of a changepoint system but simple enough to be understandable and adaptible. We focus our attention on the properties of the threshold alarm for different values of the parameters of the threshold alarm and the model. This will give us practically relevant estimates for this important class of alarm systems and moreover a benchmark for the evaluation of competing alternative algorithms. Note that virtually *every algorithm* designed to detect changepoints is based on a threshold alarm, the only difference being that the threshold alarm is not fed with the original data but by a transformation thereof, usually called "residuum" [Basseville 1993].
As a general measure for quality, we look on the one hand at the *mean delay* time τ between a changepoint and its detection and on the other hand at the mean waiting time for a false alarm, the so-called *average run length* ARL.

2 A Model for Changepoints in Biomedical Time Series

There is a huge amount of different models for changepoints and an even greater amount of variants of algorithms to detect these changepoints in noisy time series [Basseville 1993]. In most cases, one assumes a parametric setting, i.e. the time series is generated by some parametrizable probability distribution. The time at which at least one of these parameters changes is called the changepoint of the corresponding model. In such a way one models *additive* changes, e.g. changes in the mean, and *spectral* changes, e.g. changes in the variance.
Many biomedical time series can qualitatively be explained as arising from complicated and usually nonlinear internal feedback circuits [Morfill 1994], e.g. heart rate or blood pressure. The time dependence of such driven nonlinear systems can ranges from stationary, periodic, quasiperiodic to chaotic and will depend dramatically on the parameters of the special problem. However, all these time series have in common that they fluctuate about some constant mean with a corresponding variance *which is constant in time when the system is in equilibrium.* Any deviation from this "equlibrium" or "usual" behaviour is either due to internal or external disturbances of the feedback system or due to a malfunctioning of the feedback system itself. Whatever it may be, it will be reflected in the behaviour of the time series and should be detectable.
One model for a changepoint in this setting, i.e. one way to leave equilibrium, is that of adding a constant slope to the constant mean starting at the time of change. See figure 1 for an illustration and section 4 for the precise notation.

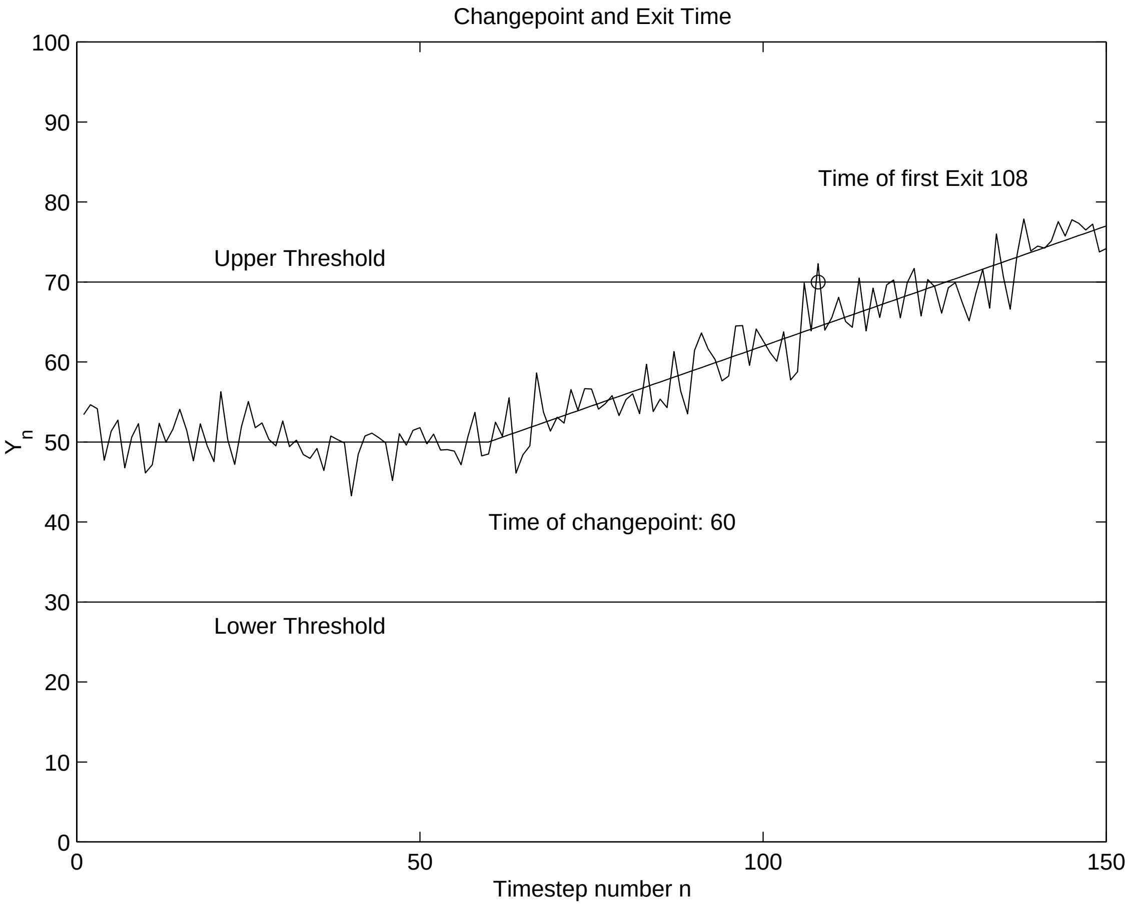

Figure 1: Simulated time series with a changepoint at time t_cp=60 and slope $\alpha = 0.3$. The series has been generated by corrupting the piecewise linear function with gaussian noise (see (1)) with zero mean $\mu_0 = 0$ and standard deviation $\sigma = 3$. The time of first exit from the threshold band between 30 and 70 is here 108. Half the width of the threshold band is $\Delta = 20$. We give robust estimates for the mean of the exit time for all α.

Even in cases where the variable of interest and its corresponding changepoints are not of this type it is frequently possible to find a transformation which casts the original problem in the setting of threshold alarms, e.g. the Fourier transformation for detection of a shift in the main frequency.

3 Heuristic Arguments for the Behaviour of ARL and Mean Delay

Before beginning with the more technical part we shall give some heuristic arguments for the mean delay and the average run length of a threshold alarm.
Looking at figure (2) we get the following impression: The average run length should increase rapidly as the threshold increases and should decrease rapidly as the variance of the noise increases. If we denote the probability for a false alarm at a single time step by $1 - p_\Delta$ we expect that on the average we should have to wait the inverse time for the next false alarm[1] i.e. the average run length should be $ARL = \frac{1}{1-p_\Delta}$, which in fact turns out to be the correct value (see equation 3).
The mean delay—for given α, which determines the average slope after the changepoint—should also increase with increasing threshold, but not as fast. We expect that the mean delay $\overline{\tau}$ should roughly be given by $\overline{\tau} \approx \frac{\Delta}{|\alpha|}$ which is certainly the zero noise limit of the delay, implying the obvious fact that small angles should have large mean delays. In the following, we shall find the quantitative justification of these statements. We will always establish robust versions of the estimates—otherwise the practical importance would be very limited.

4 Main Results

We assume that $(\varepsilon_n)_{n\in\mathbb{N}}$ is a sequence of real i.i.d. random variables defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with zero mean and finite variance, that is,

$$\mathbb{E}(\varepsilon_n) = 0 \quad \text{and} \quad V(\varepsilon_n) = \mathbb{E}(\varepsilon_n^2) = \sigma^2 > 0.$$

We now consider a second sequence $(Y_n)_{n\in\mathbb{N}}$ of random variables defined on $(\Omega, \mathcal{A}, \mathbb{P})$. This is the sequence consisting of the variables

$$Y_n \;:=\; \mu_0 + \alpha(n - t_{cp})\theta(n - t_{cp}) + \varepsilon_n, \tag{1}$$

where $\mu_0, \alpha \in \mathbb{R}$. Here, θ denotes the Thetafunction defined on the set of real numbers, that is

$$\theta(t) = \begin{cases} 1 & \text{for } t \geq 0 \\ 0 & \text{for } t < 0. \end{cases}$$

[1]There is—in general—a difference between the mean time until the next alarm ocurrs and the mean time between two alarms. See "waiting time paradoxa" in [Feller 1971].

The positive number t_{cp} is called the time of the changepoint or simply the *changepoint* of the model.
We are interested in the time of the first exit of $|Y_n(\omega)|$ from the interval $[\mu_0 - \Delta, \mu_0 + \Delta]$. For the rest of the paper we set

$$p_\Delta \quad := \quad \mathbb{P}(|\varepsilon_1| \leq \Delta) \tag{2}$$

and shall usually assume $p_\Delta < 1$ in order to exclude trivial cases.

Definition. Let $\alpha \in \mathbb{R}$.
The random variable $T_e(\alpha) : \Omega \to \mathbb{N}$ defined by

$$T_e(\alpha) := \inf\{n \in \mathbb{N} \,|\, |Y_n| > \Delta\}.$$

is called *Exit-Time.*

Remark. $RL := T_e(0)$ is also called the *Runlength* of the system and for $t_{cp} \in \mathbb{R}^+, \alpha \in \mathbb{R}\backslash\{0\}$, we call $\tau := \max(T_e - t_{cp}, 0)$ the *delay* (of detection). The *Average Runlength* $ARL(\alpha) := \mathbb{E}(T_e(\alpha))$ and the *Mean Delay* of detection $\overline{\tau}(\alpha) = \mathbb{E}(\tau(\alpha))$ (or slight modifications thereof) are the standard variables for estimating the goodness of a changepoint detection algorithm.
In the following, we may without loss of generality assume that $\mu_0 = 0$ (otherwise replace Y_n by $Y_n - \mu_0$). For $\alpha \neq 0$ we are mainly interested in the delay of detection τ and not in possible false alarms before detection and shall therefore set $t_{cp} = 0$.

We first state our main results.

Theorem 1. *Let $\alpha \in \mathbb{R}$ and p_Δ as in (2). We assume that $p_\Delta < 1$. Let the sequence $(Y_n)_{n \in \mathbb{N}}$ be defined as in (1). Then:*

$$\infty > \mathbb{E}(T_e(0)) \quad = \quad \frac{1}{1 - p_\Delta} \geq \frac{\Delta^2}{\sigma^2}, \tag{3}$$

$$V(T_e(0)) \quad = \quad \frac{p_\Delta}{(1 - p_\Delta)^2} \geq \left(\frac{\Delta}{\sigma}\right)^2 \left(\left(\frac{\Delta}{\sigma}\right)^2 - 1\right). \tag{4}$$

For $\alpha \neq 0$

$$\mathbb{E}(T_e(\alpha)) \quad \leq \quad \left(\left[2\frac{\Delta}{|\alpha|}\right] + 1\right) e. \tag{5}$$

The proof of the theorem is split into several parts.
First we determine the distribution function of $T_e(\alpha)$.

Lemma 2. *Let $T_e(\alpha)$ and Y_n be defined as above. Let $t \in \mathbb{R}, n \in \mathbb{N}$. Then*

$$\mathbb{P}(T_e(\alpha) \le t) = \begin{cases} 0 & \text{for } t < 1 \\ 1 - \prod_{i=1}^{[t]} \mathbb{P}(|Y_i| \le \Delta) & \text{for } t \ge 1 \end{cases} \tag{6}$$

and

$$\mathbb{P}(T_e(\alpha) = n) = \mathbb{P}(|Y_n| > \Delta) \prod_{i=1}^{n-1} \mathbb{P}(|Y_i| \le \Delta).$$

Proof. For $t < 1$ there is nothing to prove. From

$$\{T_e(\alpha) > n\} = \{\max(|Y_1|, \dots, |Y_n|) \le \Delta\}$$

we see that for $n \ge 1$:

$$\begin{aligned}
\mathbb{P}(T_e(\alpha) \le n) &= 1 - \mathbb{P}(T_e(\alpha) > n) \\
&= 1 - \mathbb{P}(\{|Y_i| \le \Delta \text{ for all } i = 1, \dots, n\}) \\
&= 1 - \prod_{i=1}^{n} \mathbb{P}(\{-\alpha i - \Delta \le \varepsilon_i \le -\alpha i + \Delta\}) \\
&= 1 - \prod_{i=1}^{n} \mathbb{P}(|Y_i| \le \Delta).
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathbb{P}(T_e(\alpha) = n) &= \mathbb{P}(T_e(\alpha) \le n) - \mathbb{P}(T_e(\alpha) \le n-1) \\
&= \prod_{i=1}^{n-1} \mathbb{P}(|Y_i| \le \Delta) - \prod_{i=1}^{n} \mathbb{P}(|Y_i| \le \Delta) \\
&= (1 - \mathbb{P}(|Y_n| \le \Delta)) \prod_{i=1}^{n-1} \mathbb{P}(|Y_i| \le \Delta) \\
&= \mathbb{P}(|Y_n| > \Delta)) \prod_{i=1}^{n-1} \mathbb{P}(|Y_i| \le \Delta).
\end{aligned}$$

This proves the second part of the lemma. □

Corollary 3. *For $\alpha \neq 0$*

$$\mathbb{E}(T_e(\alpha)) = 1 + \sum_{k=1}^{\infty} \prod_{i=1}^{k} \mathbb{P}(|Y_i| \le \Delta) < \infty.$$

Proof. By (6), we have

$$\begin{aligned}
\mathbb{E}(T_e(\alpha)) &= \sum_{k=0}^{\infty} \mathbb{P}(T_e(\alpha) > k) \\
&= 1 + \sum_{k=1}^{\infty} (1 - \mathbb{P}(T_e(\alpha) \leq k)) \\
&= 1 + \sum_{k=1}^{\infty} \prod_{i=1}^{k} \mathbb{P}(|Y_i| \leq \Delta).
\end{aligned}$$

We use the ratio test to prove the convergence of the inifinite series.

$$\begin{aligned}
\frac{\prod_{i=1}^{k+1} \mathbb{P}(|Y_i| \leq \Delta)}{\prod_{i=1}^{k} \mathbb{P}(|Y_i| \leq \Delta)} &= \mathbb{P}(|Y_{k+1}| \leq \Delta) \\
&= \mathbb{P}(-\alpha(k+1) - \Delta \leq \varepsilon_k \leq -\alpha(k+1) + \Delta) \\
&= \mathbb{P}(-\alpha(k+1) - \Delta \leq \varepsilon_1 \leq -\alpha(k+1) + \Delta) \\
&= \mathbb{P}(\varepsilon_1 \leq -\alpha(k+1) + \Delta) \\
&\quad -\mathbb{P}(\varepsilon_1 < -\alpha(k+1) - \Delta).
\end{aligned}$$

For $\alpha \neq 0$ the expression in the last line tends to 0 as k tends to ∞. □

Lemma 4. *Let $\alpha = 0$ and p_Δ as in (2). Then*

$$\mathbb{E}(T_e(0)) = \begin{cases} \infty & \text{for } p_\Delta = 1 \\ \frac{1}{1-p_\Delta} & \text{for } p_\Delta < 1. \end{cases}$$

Proof.

$$\begin{aligned}
\mathbb{E}(T_e(0)) &= 1 + \sum_{k=1}^{\infty} p_\Delta^k \\
&= \sum_{k=0}^{\infty} p_\Delta^k = \begin{cases} \infty & \text{for } p_\Delta = 1 \\ \frac{1}{1-p_\Delta} & \text{for } p_\Delta < 1, \end{cases}
\end{aligned}$$

as asserted. □

Using Chebyshev's inequality we immediately obtain:

Corollary 5.

$$\mathbb{E}(T_e(0)) \geq \frac{\Delta^2}{\sigma^2}.$$

Proof. For $p_\Delta = 1$, there is nothing to prove. So let $p_\Delta < 1$. Then, because of $p_\Delta \geq 1 - (\sigma/\Delta)^2$, we have

$$\begin{aligned} \mathbb{E}(T_e(0)) &= \frac{1}{1-p_\Delta} \\ &\geq \frac{\sigma^2}{\Delta^2}. \end{aligned}$$

□

Thus, (3) is proved.
We next determine the variance of $T_e(0)$.

Lemma 6. *Let $p_\Delta < 1$. Then*

$$V(T_e(0)) = \frac{p_\Delta}{(1-p_\Delta)^2}. \tag{7}$$

Proof. We have

$$\begin{aligned} \mathbb{E}(T_e(0)^2) &= \sum_{k=1}^{\infty} k^2 p_\Delta^{k-1}(1-p_\Delta) \\ &= (1-p_\Delta)\sum_{k=1}^{\infty} k(k-1+1)p_\Delta^{k-1} \\ &= p_\Delta(1-p_\Delta)\sum_{k=1}^{\infty} k(k-1)p_\Delta^{k-2} + (1-p_\Delta)\sum_{k=1}^{\infty} kp_\Delta^{k-1} \\ &= p_\Delta(1-p_\Delta)\frac{d^2}{dp_\Delta^2}\left[\frac{1}{1-p_\Delta} - 1\right] + (1-p_\Delta)\frac{d}{dp_\Delta}\left[\frac{1}{1-p_\Delta} - 1\right] \\ &= p_\Delta(1-p_\Delta)\frac{2}{(1-p_\Delta)^3} + \frac{1}{1-p_\Delta} \\ &= \frac{1+p_\Delta}{(1-p_\Delta)^2}, \end{aligned}$$

where the interchange of differentiation and summation is justified by uniform convergence. Hence,

$$\begin{aligned} V(T_e(0)) &= \mathbb{E}(T_e(0)^2) - \mathbb{E}(T_e(0))^2 \\ &= \frac{1+p_\Delta}{(1-p_\Delta)^2} - \frac{1}{(1-p_\Delta)^2} \\ &= \frac{p_\Delta}{(1-p_\Delta)^2}, \end{aligned}$$

which had to be proved. □

By Chebyshev's inequality, this gives us the following estimation of $V(T_e(0))$.

Corollary 7. *Let* $p_\Delta < 1$. *Then*

$$V(T_e(0)) \geq \left(\frac{\Delta}{\sigma}\right)^2 \left(\left(\frac{\Delta}{\sigma}\right)^2 - 1\right).$$

Proof. From (7) we obtain using $p_\Delta \geq 1 - (\sigma/\Delta)^2$:

$$\begin{aligned} V(T_e(0)) &= \frac{p_\Delta}{(1-p_\Delta)^2} \\ &\geq \left(\frac{\Delta^2}{\sigma^2}\right)^2 \left(1 - \frac{\sigma^2}{\Delta^2}\right) \\ &= \left(\frac{\Delta}{\sigma}\right)^2 \left(\left(\frac{\Delta}{\sigma}\right)^2 - 1\right). \end{aligned}$$

□

This proves (4).
We will now derive an upper bound of $\mathbb{E}(T_e(\alpha))$.

Proposition 8. *Let* $\alpha \neq 0$. *Then*

$$\mathbb{E}(T_e(\alpha)) \leq ([2\Delta/|\alpha|] + 1)e.$$

Proof. Put $\xi_n = -\varepsilon_n$, that is, $Y_n = \alpha n - \xi_n$ and $I_n = [\alpha n - \Delta, \alpha n + \Delta]$. Define

$$m := [2\Delta/|\alpha|] + 1.$$

This choice of m implies that for fixed $j \in \{1, \dots, m\}$ the intervals

$$I_1, I_{m+j}, I_{2m+j}, \dots, I_{(k-1)m+j}$$

are pairwise disjoint. Furthermore, for $k > 0$ we have

$$\{1, \dots, km\} = \bigcup_{j=1}^{m} \{1, m+j, 2m+j, \dots, (k-1)m+j\}.$$

This yields:

$$\begin{aligned} \mathbb{P}(T_e(\alpha) > km) &= \mathbb{P}(|Y_1| \leq \Delta, \dots, |Y_{km}| \leq \Delta) \\ &= \mathbb{P}(\xi_1 \in I_1, \dots, \xi_{km} \in I_{km}) \\ &= \mathbb{P}(\xi_1 \in I_1, \xi_{m+1} \in I_{m+1}, \dots, \xi_{(k-1)m+1} \in I_{(k-1)m+1}) \\ &\times \mathbb{P}(\xi_2 \in I_2, \xi_{m+2} \in I_{m+2}, \dots, \xi_{(k-1)m+2} \in I_{(k-1)m+2}) \\ &\quad \vdots \\ &\times \mathbb{P}(\xi_m \in I_m, \xi_{m+m} \in I_{m+m}, \dots, \xi_{(k-1)m+m} \in I_{(k-1)m+m}) \\ &\leq \left(\frac{1}{k!}\right)^m. \end{aligned}$$

Here, the last step follows from the following observation: Since for $j \in \{1, \dots, m\}$ the intervals $I_j, I_{m+j}, \dots, I_{(k-1)m+j}$ are pairwise disjoint, we have

$$\begin{aligned} & \mathbb{P}(\xi_j \in I_j, \xi_{m+j} \in I_{m+j}, \dots, \xi_{(k-1)m+j} \in I_{(k-1)m+j}) \\ \leq\ & \mathbb{P}(\xi_j < \xi_{m+j} < \dots < \xi_{(k-1)m+j}) \\ \leq\ & \frac{1}{k!}. \end{aligned}$$

The general inequality now follows since each ordering of $\xi_j, \xi_{m+j}, \dots, \xi_{(k-1)m+j}$ has the same probability because the sequence $(\xi_n)_{n\in\mathbb{N}}$ is identically distributed. We now obtain

$$\begin{aligned} \mathbb{E}(T_e(\alpha)) &= \sum_{n=0}^{\infty} \mathbb{P}(T_e(\alpha) > n) \\ &= \sum_{k=0}^{\infty} \sum_{l=0}^{m-1} \mathbb{P}(T_e(\alpha) > km + l) \\ &\leq \sum_{k=0}^{\infty} m\mathbb{P}(T_e(\alpha) > km) \\ &\leq m \sum_{k=0}^{\infty} \left(\frac{1}{k!}\right)^m \\ &\leq ([2\Delta/|\alpha|] + 1)e, \end{aligned}$$

which is (5). $\square$

5 Conclusion

We have studied the perhaps most frequently used alarm system for changepoint detection—the threshold alarm—for a model which seems to cover many types of time series appearing in biomedical applications. We suceeded to derive robust estimates for the average run length (ARL) and the mean delay for detection which are the most important criteria for the quality of a changepoint detection algorithm.

What is astonishing are the mathematical difficulties in justifying obvious conjectures about the behaviour of ARL and especially mean delay even in the simple case of threshold alarms. From a mathematical point of view it would be desirable to obtain analytic control of the complete ARL function, i.e. the behaviour of the run length in dependence of the slope α which at present we only know to do via simulations. Moreover, we would like to have such a good control also for more complex adaptive algorithms and also for extensions of our model which includes sequential changepoints. Simulations will certainly help to design appropriate algorithms but any analytic result in this context will cut down the necessary computational power siginifantly and moreover serve as valuable cross-check for plausibility of the results. We shall report about such adaptive algorithms and sequential changepoints elsewhere. From a practical point of view we have now sufficiently good control of the last (and sometimes only) stage of any changepoint detection algorithm. For some medical applications see [Daumer 1997].

References

[Basseville 1993] Basseville, M., Nikiforov, I.V., *Detection of Abrupt Changes*, Prentice Hall, Englewood Cliffs, New Jersey, 1993.

[Daumer 1997] Daumer, M., Online Monitoring von Changepoints, to appear in: Ergänzungsband Biomedizinische Technik, Verlag Schiele und Schön (1997)

[Feller 1971] Feller, W., An Introduction to Probability Theory and Its Applications, John Wiley & Sons, New York-London-Sydney-Toronto, Volume II (1971)

[Morfill 1994] Morfill, G., Schmidt, G. Komplexitätsanalyse in der Kardiologie, *Physikalische Blätter* **50** (1994), 156-160.

[Smith 1983] Smith, A.F.M. and West, M., Monitoring Renal Transplants: An application of the Multiprocess Kalman Filter, *Biometrics* **39** (1983), 867–878.